

Little Mathematics Library



L. A. LYUSTERNIK

THE
SHORTEST
LINES
VARIATIONAL
PROBLEMS

Mir Publishers • Moscow



ПОПУЛЯРНЫЕ ЛЕКЦИИ ПО МАТЕМАТИКЕ

Л. А. Люстерник

КРАТЧАЙШИЕ ЛИНИИ

ВАРИАЦИОННЫЕ ЗАДАЧИ

**ИЗДАТЕЛЬСТВО «НАУКА»
МОСКВА**

LITTLE MATHEMATICS LIBRARY

L.A. Lyusternik

THE
SHORTEST
LINES

VARIATIONAL
PROBLEMS

Translated from the Russian
by
Yuri Ermolyev

.

MIR PUBLISHERS
MOSCOW

TO THE READER

Mir Publishers would be grateful for your comments on the content, translation and design of this book. We would also be pleased to receive any other suggestions you may wish to make.

Our address is:

Mir Publishers
2 Pervy Rizhsky Pereulok
I-110, GSP, Moscow, 129820
USSR

First published 1976

Second printing 1983

На английском языке

© English translation, Mir Publishers, 1976

CONTENTS

Introduction	7
------------------------	---

LECTURE 1

Chapter I. Shortest Lines on Simple Surfaces	9
I.1. Shortest Lines on Polyhedral Surfaces	9
I.2. Shortest Lines on a Cylindrical Surface	14
I.3. Shortest Lines on a Conical Surface	22
I.4. Shortest Lines on a Spherical Surface	30

Chapter II. Some Properties of Plane and Space Curves and Associated Problems	38
--	----

II.1. Tangent and Normals to Plane Curves and Associated Problems	38
II.2. Some Information on the Theory of Plane and Space Curves	42
II.3. Some Information on the Theory of Surfaces	46

Chapter III. Geodesic Lines	49
---------------------------------------	----

III.1. Bernoulli's Theorem on Geodesic Lines	49
III.2. Additional Remarks Concerning Geodesic Lines	54
III.3. Geodesic Lines on Surfaces of Revolution	58

LECTURE 2

Chapter IV. Problems Associated with the Potential Energy of Stretched Threads	61
IV.1. Motion of Lines that Does Not Change Their Length . .	61
IV.2. Evolutes and Involutives	67
IV.3. Problems of Equilibrium of a System of Elastic Threads	68
 Chapter V. The Isoperimetric Problem	 73
V.1. Curvature and Geodesic Curvature	73
V.2. An Isoperimetric Problem	76
 Chapter VI. Fermat's Principle and Its Corollaries	 82
VI.1. Fermat's Principle	82
VI.2. The Refraction Curve	85
VI.3. The Problem of Brachistochrone	89
VI.4. The Catenary and the Problem of the Minimal Surface of Revolution	92
VI.5. Interrelation Between Mechanics and Optics	101

INTRODUCTION

This book is an attempt to examine from the elementary point of view a number of so-called variational problems. These problems deal with quantities dependent on a curve, and a curve for which this quantity is either maximum or minimum is sought. Such are, for example, problems in which it is required to find the shortest of all the curves connecting two points on a surface, or among all the closed curves of a given length on a plane it is necessary to find that one which bounds the maximum area, and so on.

The material of this book was basically presented by the author in his lectures at a school mathematical circle of the Moscow State University. The contents of the first lecture (Secs. I.1-III.3) in the main coincides with the contents of *Geodesic Lines*, published by the author in 1940.

Only the knowledge of elementary mathematics is required to master this course. Moreover, the first chapters are quite elementary. Others while not requiring special knowledge demand a greater aptitude for mathematical perusal and meditation.

The entire material of this book may be considered as an elementary introduction to the calculus of variations (a branch of mathematics dealing with problems of finding the functional minimum or maximum). The calculus of variations does not enter into the first concentric cycle of the higher mathematics course that is studied, for example, in technical colleges. In our opinion, however, for a student who begins to study higher mathematics, it is not useless to look further ahead.

For a reader familiar with elements of mathematical analysis it will not be difficult to make certain definitions and reasonings (not strictly stated in this book) absolutely

strict (appropriate explanatory considerations will be often found in the text in small print). For example, we should speak not about small quantities and their approximate equality, but about infinitesimals and their equivalence. If, however, a more exacting reader remains unsatisfied with the level of rigidity and logical completeness accepted here, let this be used for explaining the necessity of such logical polishing of the basic ideas of the mathematical analysis which will be encountered, for example, in the university courses. Without this polishing a strict and systematic presentation of such chapters of analysis as the calculus of variations is impossible.

Mathematical analysis has worked out a powerful analytical apparatus sometimes capable of solving many difficult questions. However, at all stages of studying mathematics it is of primary importance to see the simple geometric or physical sense of the problem. One should be able to give rough trial solutions to problems in a simple, non-rigorous fashion.

If this small book helps, to a certain degree, to develop those elements of mathematical culture in the reader, the author will consider the work spent on writing it not wasted.

LECTURE 1

CHAPTER I

Shortest Lines on Simple Surfaces

I.1. Shortest Lines on Polyhedral Surfaces

1. The shortest line on a dihedral angle. Everybody knows that a straight-line segment is the shortest of all the lines connecting two points on a plane.

Consider two points A and B on an arbitrary surface. They may be connected by an infinite number of various lines lying on the surface. Which of those lines is the shortest? In other words, how should one move about the surface in order to get from point A to point B by the shortest route?

First let us solve this problem for some elementary surfaces. We shall begin with the following problem. Consider a dihedral angle* with faces Q_1 and Q_2 and edge MN (Fig. I.1). Given two points: point A on Q_1 and point B on Q_2 . Points A and B can be connected by an infinite number of different lines lying on faces Q_1 and Q_2 of the dihedral angle. Find the shortest of the lines.

If the dihedral angle is equal to two right angles (180°), faces Q_1 and Q_2 will be a continuation of each other, i.e. will make one plane, and straight-line segment AB connecting points A and B is the shortest line sought for. If, on the other hand, the dihedral angle is not equal to two right angles, faces Q_1 and Q_2 will not be a continuation of each other and straight-line segment AB will not lie on these faces. Let us turn one of the faces about straight line MN so that the faces become a continuation of each other, in other words, unfold the dihedral angle onto a plane (Fig. I.2).

* Only part of this infinite dihedral angle is represented in Fig. I.1.

Then faces Q_1 and Q_2 will pass into half-planes Q'_1 and Q'_2 . Straight line MN will turn into straight line $M'N'$ separating Q'_1 and Q'_2 ; points A and B will become points A' and B' (A' located on Q'_1 and B' on Q'_2). Each line lying on the faces of the dihedral angle and connecting points A and B will become a line of the same length connecting points A' and B' on this plane. The shortest line on the faces of the dihedral angle, i.e. the line connecting points A and B , will become the shortest line connecting points A' and B' on the plane, i.e. will pass into straight-line segment $A'B'$. This segment will intersect straight line $M'N'$ at a

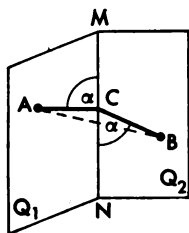


Fig. I.1

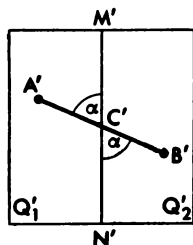


Fig. I.2

certain point C' ; angles $A'C'M'$ and $N'C'B'$ will be equal to each other as the vertically opposite ones (see Fig. I.2). Denote each of them by α .

Now turn Q'_1 and Q'_2 about $M'N'$ so as to again obtain the primary dihedral angle. Half-planes Q'_1 and Q'_2 will again become faces Q_1 and Q_2 of this dihedral angle, $M'N'$ its edge MN , and points A' and B' points A (on face Q_1) and B (on face Q_2), straight-line segment $A'B'$ will pass into the shortest line lying on the faces of the dihedral angle, i.e. the line connecting points A and B . Obviously, this shortest line is broken line ACB , where segment AC is located on face Q_1 , and CB on face Q_2 . It is easily seen that angles ACM and NCB , into which angles $A'C'M'$ and $N'C'B'$ have turned, are equal to α as before, and, consequently, are equal to each other. Thus *the shortest of the lines lying on the faces of a dihedral angle and connecting two points A and B thereof (points located on its different faces) is broken line ACB with its vertex C on edge MN ; angles ACM and NCB , formed by the segments of the broken line with the edge, are equal to each other.*

The problem is sometimes livened up. A fly intends to crawl from point A situated on one wall to point B situated on the adjacent wall. What is the shortest route for it to get from point A to point B ? After the above discussion the problem presents no difficulty.

2. The shortest line on a polyhedral surface. Now let us consider a more complicated case. Given a polyhedral surface (Fig. I.3) consisting of several faces $Q_1, Q_2, Q_3, Q_4, \dots$

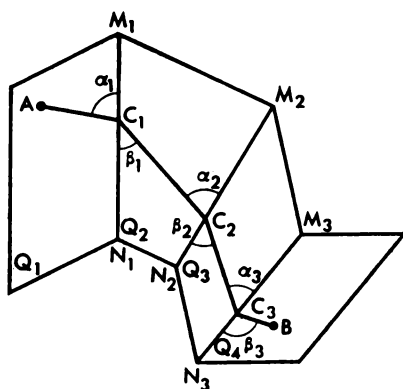


Fig. I.3

\dots, Q_n with edges $M_1N_1, M_2N_2, \dots, M_{n-1}N_{n-1}$ (in Fig. I.3 $n = 4$). Points A and B lie on two different faces of the polyhedral surface (for example, on Q_1 and Q_4). Find the shortest line connecting points A and B .

Let this line AB be the shortest one and let it pass over faces Q_1, Q_2, Q_3 and Q_4 . Unfold part of the polyhedral surface consisting of these faces onto a plane (Fig. I.4). This being done, the faces will pass into polygons Q'_1, Q'_2, Q'_3 and Q'_4 , and edges M_1N_1, M_2N_2 and M_3N_3 which adjoined faces Q_1, Q_2, Q_3 and Q_4 will turn into sides $M'_1N'_1, M'_2N'_2$ and $M'_3N'_3$ of polygons Q'_1, Q'_2, Q'_3 and Q'_4 along which the latter adjoin each other. Points A and B will pass into points A' and B' of the plane, and the lines connecting them on the folded surface will pass into lines on the plane, i.e. the lines connecting A' and B' . The shortest of the lines connecting points A and B will become the shortest plane line connecting points

A' and B' , i.e. straight-line segment $A'B'$ *. Here the previous reasoning will be fully applicable, viz. vertical angles

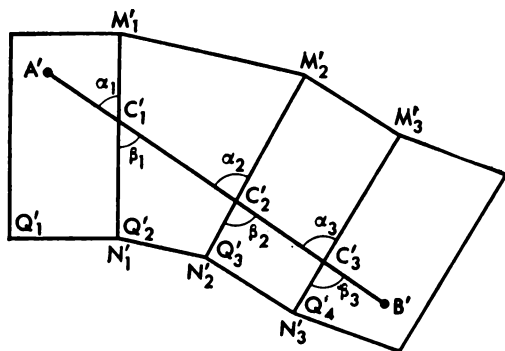


Fig. I.4

α_1 and β_1 formed by straight-line segment $A'B'$ with side M_1N_1 are equal. In a similar manner vertical angles α_2 and β_2 , α_3 and β_3 , formed by straight-line segment $A'B'$ with sides M_2N_2 and M_3N_3 (see Fig. I.4), are pairwise equal.

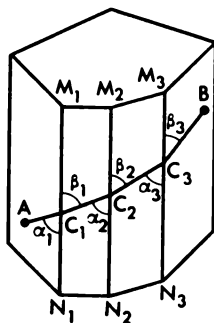


Fig. I.5

If we again fold part of the plane made up of those polygons into a polyhedral surface so that polygon Q_1 again becomes face Q_1 , polygon Q_2 face Q_2 , polygon Q_3 face Q_3 and, finally, polygon Q_4 face Q_4 , points A' and B' will turn into points A and B , and segment $A'B'$ will become line \widehat{AB} , i.e. the shortest of all the lines on the polyhedral surface connecting points A and B . This shortest line will be a broken line whose vertices lie on edges M_1N_1 , M_2N_2 and M_3N_3 of the

polyhedral surface. Angles α_1 and β_1 (just as α_2 and β_2 , α_3 and β_3), which form together with the surface edge the two adjacent segments of the broken line, are equal.

3. The shortest line on the lateral surface of a prism. Figure I.5 represents a prism** with the shortest line on its

* We do not consider here the case when $A'B'$ intersects other sides of these polygons.

** The faces of the prism are assumed to be unbounded.

surface connecting two points A and B lying on different faces of the prism. This shortest line is a broken line with vertices C_1 , C_2 and C_3 on the edges of the prism. The angles formed by the two adjacent segments of the line together with the edge of the prism, on which their common vertex lies, are equal, exactly as in the previous case:

$$\alpha_1 = \beta_1, \quad \alpha_2 = \beta_2, \quad \alpha_3 = \beta_3, \quad \dots$$

But besides that we have $\beta_1 = \alpha_2$.

Indeed, these are alternate interior angles with parallel straight lines M_1N_1 , M_2N_2 and secant C_1C_2 . Similarly $\beta_2 = \alpha_3$.

Thus we have

$$\alpha_1 = \beta_1 = \alpha_2 = \beta_2 = \alpha_3 = \beta_3 = \dots$$

In other words, *the angles formed by the segments of the shortest broken line AB on the surface of a prism with all the edges are equal.*

4. The shortest line on the surface of a pyramid. Given two points A and B on the lateral faces of a pyramid* with vertex O (Fig. I.6). These points may be connected on the surface of the pyramid by an infinite number of lines, one of which

\widehat{AB} is the shortest. On the basis of the previous reasoning, \widehat{AB} is a broken line whose vertices C_1 , C_2 , C_3, \dots lie on the edges of the pyramid, and angles α_1 and β_1 , α_2 and β_2 , α_3 and β_3, \dots , formed by the segments of the broken line with the edges of the pyramid, are pairwise equal:

$$\alpha_1 = \beta_1, \quad \alpha_2 = \beta_2, \quad \alpha_3 = \beta_3, \quad \dots$$

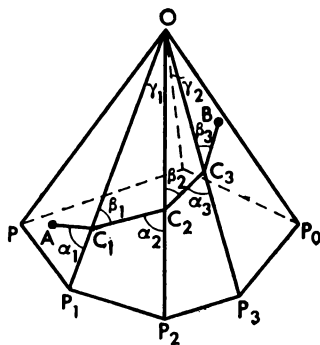


Fig. I.6

Consider face P_1OP_2 and segment C_1C_2 on it. If γ_1 denotes angle P_1OP_2 , then in triangle C_1OC_2 angle α_2 is exterior and angles β_1 and γ_1 are interior. The exterior angle of a triangle is equal to the sum of the interior angles not adja-

* Its faces are assumed to be unlimitedly extended.

cent to it; hence,

$$\alpha_2 = \beta_1 + \gamma_1, \quad \text{or} \quad \alpha_2 - \beta_1 = \gamma_1$$

But since $\beta_1 = \alpha_1$, $\alpha_2 - \alpha_1 = \gamma_1$.

Similarly $\alpha_3 - \alpha_2 = \gamma_2$, where γ_2 is the angle at vertex O between two neighbouring edges OP_2 and OP_3 , etc.

Thus *the difference in the angles at which the shortest line intersects any two edges of a pyramid is equal to the sum of the corresponding plane angles at the vertex.*

I.2. Shortest Lines on a Cylindrical Surface

1. The shortest line on a cylindrical surface. Now let us consider the method of finding the shortest lines on certain

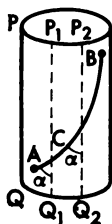


Fig. I.7

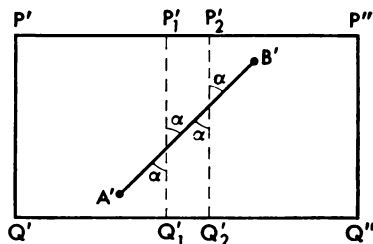


Fig. I.8

simple curved surfaces. We shall start with the surface of a circular cylinder*.

It will be recalled that a cylindrical surface can be covered by a system of straight lines parallel to the axis of the cylinder and, hence, to one another. These straight lines are called the *generatrices of a cylinder*.

Consider two points A and B on the surface of a cylinder (Fig. I.7). Find the shortest of the curves lying on the cylindrical surface and connecting points A and B . Denote this shortest line connecting points A and B by \widehat{AB} . First consider the case when A and B do not lie on the same generatrix.

Cut the lateral surface of the cylinder along a certain generatrix PQ (not intersecting \widehat{AB}) and develop a plane. We shall obtain a rectangle (Fig. I.8) (one pair of the sides, $P'P''$ and $Q'Q''$, is due to developing the circumferences

* The surface of the finite cylinder under consideration (Fig. I.7) is part of the surface of an infinite cylinder.

bounding the lateral surface of the cylinder, and the other pair, $P'Q'$ and $P''Q''$, is formed by two edges of section PQ). The generatrices of the cylinder will turn into straight lines parallel to side $P'Q'$ of the rectangle. Points A and B will turn into points A' and B' lying inside the rectangle. Lines connecting points A and B on the cylinder become straight lines connecting points A' and B' inside the rectangle. Arc \widehat{AB} , i.e. the shortest of the lines on the cylinder connecting points A and B , will pass into the shortest straight line connecting points A' and B' , i.e. into straight-line segment $A'B'$. Thus after developing the lateral surface of the cylinder into a plane rectangle *the shortest arc \widehat{AB} on the cylindrical surface turns into straight-line segment $A'B'$* . Generatrices P_1Q_1, P_2Q_2, \dots of the cylinder will turn into straight lines $P'_1Q'_1, P'_2Q'_2, \dots$ parallel to sides $P'Q', P''Q''$ of rectangle $P'Q'Q''P''$. The angles formed by segment $A'B'$ with these straight lines are equal, as the corresponding angles at parallel lines. Denote each of them by α .

Fold rectangle $P'Q'Q''P''$ (gluing together its opposite sides $P'Q'$ and $P''Q''$) so that it again assumes the original cylindrical shape. Points A' and B' will again pass into points A and B of the cylinder, and straight-line segment $A'B'$ connecting them into the shortest arc \widehat{AB} on the cylindrical surface. The angles of segment $A'B'$ with straight lines $P'_1Q'_1, P'_2Q'_2$ will turn into angles between arc \widehat{AB} and generatrices P_1Q_1, P_2Q_2, \dots equal to the former. Since straight-line segment $A'B'$ intersects all the straight lines parallel to $P'Q'$ at equal angles α , *the shortest arc \widehat{AB} into which $A'B'$ passes will intersect all the generatrices of the cylinder at equal angles α* (see Fig. I.7).

Consider a special case when points A and B are on the same generatrix (Fig. I.9). It is evident that *segment AB of the generatrix will be the shortest distance between points A and B on the cylindrical surface*.

One more case may be singled out where points A and B are on one circular section of the cylinder (Fig. I.10). Arc \widehat{AB} of this section is normal to all the generatrices and is the shortest arc connecting points A and B .

If a cylinder is cut along the generatrix not intersecting arc \widehat{AB} and is developed into a plane rectangle, then in the

two singled-out cases the shortest arc will pass into a segment parallel to the sides of the rectangle. In all other cases

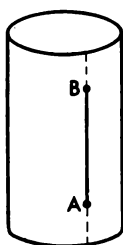


Fig. 1.9

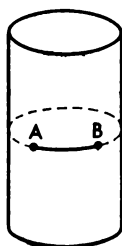


Fig. 1.10

the shortest line will intersect the generatrices at an angle different from the right angle (and not equal to zero).*

2. Helixes. A *helix* is such a line on a cylindrical surface which intersects all the generatrices of the cylinder at equal angles different from the right angle.

Denote the angle between a helix and a generatrix by α . The line intersecting the generatrices of a cylinder at right angles is the circular section. The circular section may be considered the limiting case of a helix when α is turned into the right angle. Similarly the generatrix of a cylinder may be considered as another limiting case when α is turned into zero.

Let us consider the motion over the surface of a cylinder (1) parallel to the axis (along the generatrix) and (2) in rotation about the axis (along the circular section) at a constant speed.

Each of these motions may proceed in two opposite directions. Let us call the upward motion on the vertical cylinder positive and the downward motion negative. Let us call the rotation on the vertical cylinder from right to left (as seen by a person standing upright along the axis) positive or counterclockwise, and the rotation from left to right negative or clockwise.

The motion along a helix is obtained by adding two motions: one parallel to the axis of a cylinder and the other in

* It is interesting to compare our problem of finding the shortest line on a cylindrical surface with the problem on pp. 12-13 of finding the shortest broken line on the surface of a prism (for which our problem is the limiting case).

rotation about the axis. A helix is called *right-handed* if the upward motion combines with the positive rotation (from

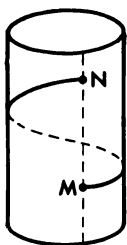


Fig. I.11

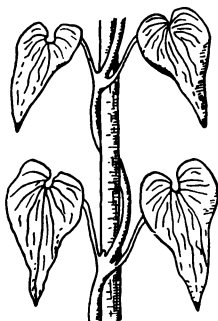


Fig. I.12

right to left) (Fig. I.11) and *left-handed* if the upward motion combines with the negative rotation (from left to right).

Most creepers (bindweed, bean) waving about the vertical support take the shape of right-handed helices (Fig. I.12). Hop, on the other hand, takes the shape of the left-handed helix (Fig. I.13).

Let a point moving along a helix intersect a generatrix at point M . Continuing the motion along the helix it will again intersect the same generatrix at point N . When the point passes

arc \widehat{MN} of the helix, it describes a circle around the cylinder's axis. At the same time it rises to a distance of the length of section MN (see Fig. I.11). The first limiting case occurs when the speed of the rotary motion is zero and the point moves parallel to the axis of the cylinder along the generatrix. The



Fig. I.13

second limiting case occurs when the speed of motion parallel to the axis of the cylinder is zero and the point simply rotates about the axis along a circumference.

From this follows the

Theorem. *The shortest arc \widehat{AB} on the cylindrical surface connecting two given points A and B is the arc of a helix.*

3. The arcs of helixes connecting two given points. Two points on a cylindrical surface can be connected by different arcs of helixes. Indeed, let two points on the cylindrical surface be connected by the shortest arc \widehat{AB} , the latter being the arc of a helix. In the case of developing the cylindrical surface (cut along a generatrix not intersecting arc \widehat{AB}) into a plane rectangle, the arc will pass into a straight-line segment (see Figs. I.7 and I.8).

Now cut the cylinder along generatrix P_1Q_1 which intersects the shortest arc \widehat{AB} at point C (see Fig. I.7). Line \widehat{AB} appears to be cut into two parts: \widehat{AC} and \widehat{CB} . If the surface of the cylinder is developed into a plane rectangle, points A and B will turn into points A'' and B'' of the rectangle (Fig. I.14), and sections \widehat{AC} and \widehat{CB} of arc \widehat{AB} will respectively turn into straight-line segments $A''C''$ and $B''C'$. But points A'' and B'' may be connected by straight-line segment $A''B''$ lying inside rectangle $P_1'Q_1'Q_1''P_1''$. Obviously, $A''B''$ is shorter than any other line connecting points A'' and B'' and lying inside this rectangle.

Again roll up the rectangle into a cylinder, gluing the lateral sides $P_1'Q_1'$ and $P_1''Q_1''$ so that point C' merges with point C'' and assumes position C . Then points A'' and B'' will again pass into points A and B on the cylindrical surface, and segments $A''C''$ and $B''C'$ will turn into arc \widehat{AB} , i.e. the shortest of all the lines connecting points A and B on the cylindrical surface, whereas segment $A''B''$ will turn into an arc of the helix AB connecting the same points A and B . In Fig. I.15 arc \widehat{AB} is the right-handed and AB the left-handed helix passing through points A and B .

Lines which do not intersect the sides of a rectangle after it has been rolled up into a cylinder will pass into lines not intersecting generatrix P_1Q_1 (since sides $P_1'Q_1'$ and $P_1''Q_1''$ of the rectangle were glued along this line). The shortest of these lines is the arc $AB = \widehat{AmB}$ (see Fig. I.15). But it may not prove to be the shortest of all the lines connecting points A and B on the surface of the cylinder because if \widehat{AB}

is shorter than \widehat{AB} , then \widehat{AB} is not the shortest of the curves lying on the surface of the cylinder and connecting points A and B .

Draw half-plane R_1 through point A and the axis of a cylinder, and half-plane R_2 through point B and the axis of the cylinder (see Fig. I.15).

These half-planes form two dihedral angles. One of them includes arc \widehat{AB} and the other arc \widehat{AB} , the shorter of them lying inside the smaller dihedral angle.

If on the other hand half-planes R_1 and R_2 are a continuation of each other (i.e. the angle between them is equal to

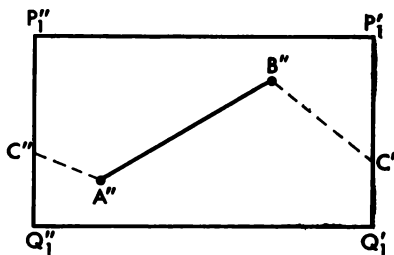


Fig. I.14

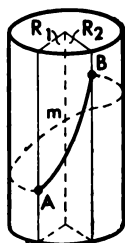


Fig. I.15

two right angles), both arcs \widehat{AB} and \widehat{AB} are equal in length. In this case there are two shortest arcs on the cylindrical surface (arcs of equal length) connecting points A and B (Fig. I.16).

The two helixes \widehat{AB} and \widehat{AB} connecting points A and B possess one property in common: moving along one of them from point A to point B we shall not describe a circle around the axis of the cylinder.

Let a rectangular sheet of paper (with a width equal to the height of the cylinder) be repeatedly wrapped around the cylinder (Fig. I.17). Pierce this sheet with a needle in points A and B and then develop it into a plane rectangle. The pinholes will be found in several points of the sheet. In Fig. I.18 they are denoted A'_1, A'_2, A'_3, \dots . These traces lie on one horizontal straight line parallel to the sides of the rectangle. If we draw straight lines $P'_1Q'_1, P'_2Q'_2, P'_3Q'_3, \dots$ parallel to the other pair of the sides of the rectangle through

points A'_1, A'_2, A'_3, \dots , they will cut off rectangle $P'_1Q'_1Q'_2P'_2$ describing one circle of the sheet around the cylinder. In wrapping the sheet around the cylinder, segments $P'_1Q'_1$

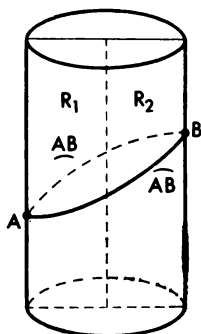


Fig. I.16

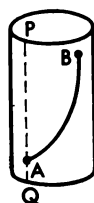


Fig. I.17

and $P'_2Q'_2$ will coincide with generatrix PQ of the cylinder passing through point A ; the merged points A'_1, A'_2 will coincide with point A of the cylinder.

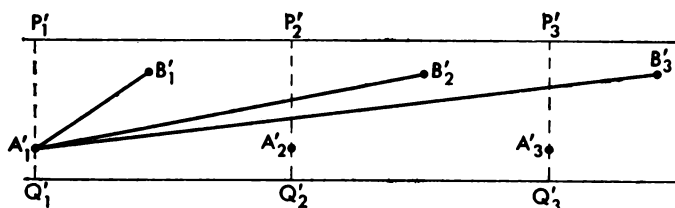


Fig. I.18

Points B'_1, B'_2, B'_3, \dots on the sheet are the traces of piercing point B of the cylinder. Their arrangement is similar to that of points A'_1, A'_2, A'_3, \dots .

Connect point A'_1 by straight lines with points B'_1, B'_2, B'_3, \dots . Again wrap the sheet around the cylinder so that points A'_1, A'_2, A'_3, \dots lie on point A and points B'_1, B'_2, B'_3, \dots on point B of the cylinder. Straight-line segment $A'_1B'_1$ will turn into arc \widehat{AB} of the helix (see Fig. I.17) discussed above.

For the sake of brevity let us say that curve AB describes n whole positive (negative) circles around the axis of the cylinder.

der if, while moving along this curve on the cylindrical surface from point A to point B , we shall describe more than n and less than $(n + 1)$ positive (negative) circles around the axis of the cylinder, or maybe exactly n circles.

When wrapping the plane around the cylinder, segment A_1B_2 will also pass into the arc of a helix $(\widehat{AB})_1$ connecting points A and B (Fig. I.19); similarly segments A_1B_3 , A_1B_4 , ...

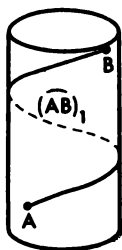


Fig. I.19

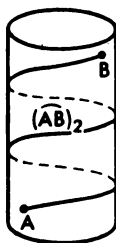


Fig. I.20

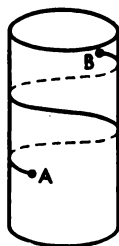


Fig. I.21

will pass into the arcs of helices $(\widehat{AB})_2$ (Fig. I.20), $(\widehat{AB})_3$, ... connecting these points. Arc $(\widehat{AB})_1$ describes one positive circle around the axis of the cylinder, arcs $(\widehat{AB})_2$, $(\widehat{AB})_3$, etc. describe two, three, etc. such circles, respectively.

Arc $(\widehat{AB})_1$ is the shortest among those connecting points A and B and describing one positive circle around the axis. Similarly $(\widehat{AB})_2$, $(\widehat{AB})_3$, etc. are the shortest of the arcs describing two, three, etc. circles, respectively.

The considered arcs were those of *right-handed* helices. Similarly arcs of *left-handed* helices may be obtained that connect points A and B and describe one, two, three, etc. negative circles around the axis of the cylinder (Fig. I.21). Each of these arcs is the shortest of the lines connecting points A and B and describing the respective number of negative circles around the axis of the cylinder.

Let us see the arrangement on the cylindrical surface of a tightly stretched rubber thread fastened at points A and B . While stretching, this line will occupy one of the shortest lines, i.e. it will lie along one of the helices connecting points A and B . If, for example, the thread is wound around the cylinder in such a way that when moving along it a posi-

tive rotation about the axis (from right to left) is to be accomplished, then the thread will occupy the position of one of the helixes \widehat{AB} , $(\widehat{AB})_1$, $(\widehat{AB})_2$, More precisely, it will occupy position \widehat{AB} if the thread does not make a single circle around the axis of the cylinder; $(\widehat{AB})_1$ if it makes one such circle, $(\widehat{AB})_2$ if it makes two circles, etc.

Indeed, on a plane rectangle the thread stretched between point A'_1 and anyone of the points B'_1 , B'_2 , B'_3 , . . . will lie along one of the segments $A'_1B'_1$, $A'_1B'_2$, $A'_1B'_3$, If we roll this sheet onto a cylindrical surface so that A'_1 merges with point A while points B'_1 , B'_2 and B'_3 with point B , the stretched thread will assume the shape of one of the helixes \widehat{AB} , $(\widehat{AB})_1$, $(\widehat{AB})_2$,

I.3. Shortest Lines on a Conical Surface

1. The shortest line on a conical surface. Let two infinite rays OA and ON come out of point O . Rotate ray OA about ray ON . The surface circumscribed in the process is called the *conical surface* (Fig. I.22); ON is called the *axis of the cone*. The rays coming out of point O and lying on the conical surface are called the *generatrices*.*

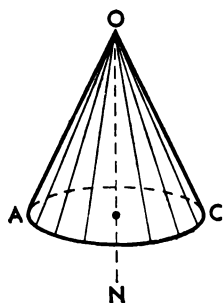


Fig. I.22

If the plane passing through generatrices OA and OC also passes through the axis of the cone, then the generatrices are called *opposite*. Two opposite generatrices divide the cone into two *equal (congruent) parts*. Cut a conical surface along generatrix OA ; after this develop the conical surface on a plane. Vertex O of the cone will become point O' on the plane; the generatrices of the cone will turn into rays coming out of O' on the plane. The entire conical surface will become angle $A'_1O'A'_2$ of the plane (Fig. I.23). The angle whose magnitude is always less than 360° is termed the *developed cone angle*. The legs of the angle $O'A'_1$ and $O'A'_2$ are made from generatrix OA along

* Only part of an infinite cone is represented in Fig. I.22.

which the conical surface was cut. Generatrix OC opposite to generatrix OA will pass into bisector $O'C'$ of angle $A_1'O'A_2'$. Indeed, both generatrices OA and OC divide the conical surface, cut along OA , into two equal parts, S and T . When this surface is developed into a plane angle $A_1'O'A_2'$, each of the parts S and T of the cone becomes halves S' and T' of this angle, and generatrix OC becomes the bisector $O'C'$ of this angle.

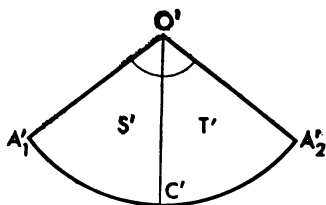


Fig. 1.23

We were developing a cut conical surface onto a plane. Now let us conduct an opposite operation, viz. rolling up angle $A_1'O'A_2'$ into a cone. Point O' will turn into vertex O of the cone, and legs $O'A'_1$ and $O'A'_2$ into the same generatrix.

Cut the plane along side $O'A'_1$ of this angle. Roll the cut plane around the cone. Generally speaking, this plane will cover the cone several times. For example, if the developed cone angle is 90° , the plane will cover the conical surface four times. Precisely, if rays $O'A'_2$, $O'A'_3$ and $O'A'_4$ are drawn from point O' at angles 90° , 180° and 270° to $O'A'_1$, then in rolling the cut surface onto the cone each of the angles $A_1'O'A'_2$, $A_2'O'A'_3$, $A_3'O'A'_4$ and $A_4'O'A'_1$ will fully cover the conical surface. Altogether the cone will be covered four times. Rays $O'A'_1$, $O'A'_2$, $O'A'_3$ and $O'A'_4$ of the plane will pass into one and the same generatrix of the cone.

If the developed cone angle is, say, 100° , the cut surface will fully overlap the conical surface threefold, and in addition a part of the cone will be covered a fourth time (the plane consists of three adjacent angles of 100° with the vertex in O' , each overlapping the entire conical surface once, and of an angle of 60° which will additionally cover part of this surface).

2. Geodesic lines on a conical surface. Let us consider an arbitrary straight line l' on a plane. Let it pass through point O' . Hence, it consists of two rays $O'D'$ and $O'E'$ (Fig. 1.24). In wrapping the surface on the cone (when point O' merges with vertex O of the cone) each of the rays $O'D'$

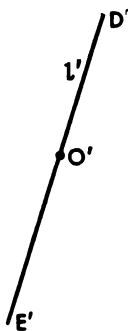


Fig. 1.24

and $O'E'$ will turn into a generatrix of the cone. Our straight line will pass into two generatrices.*

Suppose now that straight line l' does not pass through point O' (Fig. I.25). Cut the plane along ray $O'A'$ parallel to l' and wrap the cut plane around the conical surface. Straight line l' will pass into curve l on the conical surface (Fig. I.26). This curve l is called a *geodesic line* on the conical

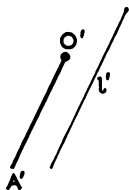


Fig. I.25

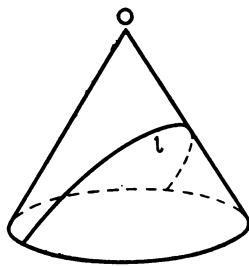


Fig. I.26

surface. Each segment of straight line l' will turn into an arc of curve l . Vice versa, each arc of curve l will turn into a segment of straight line l' when the conical surface is developed on a plane.

The obtained curves on the conical surface are similar in their function to helixes on the surface of a cylinder.

Connect points A and B on the conical surface by various lines lying on the surface and let one of them, say arc \widehat{AB} , be the shortest. When developing the surface of the cone on the plane, arc \widehat{AB} will turn into a plane arc $A'B'$. Since arc \widehat{AB} is the shortest of the lines connecting A and B and lying on the conical surface, $A'B'$ is the shortest of the lines connecting A' and B' on the plane. Hence, $A'B'$ is a straight-line segment. Arc \widehat{AB} which passes into a straight-line segment when developing the conical surface on the plane is a geodesic arc.

* Two generatrices may merge into one. This occurs when the numerical value of the developed cone angle expressed in degrees is the divisor of number 180, i.e., if the angle is equal to 180° , 90° , 60° , ..., in general, to $180^\circ/k$, where k is an integer.

As will be seen, the shape of a geodesic line depends to a great degree on the developed cone angle.

3. Double points of geodesic lines. First let us introduce the following definition. Moving along line q we may pass point A twice. Point A is called the *double point* of line q .* In Fig. I.27 point B is the double point of line l , since moving along line l in the direction shown by the arrows one will pass point B twice.

Theorem 1. If the developed cone angle is greater than, or equal to, 180° , geodesic lines on it have no double points. If, on the other hand, the angle is less than 180° , each geodesic line has at least one double point.

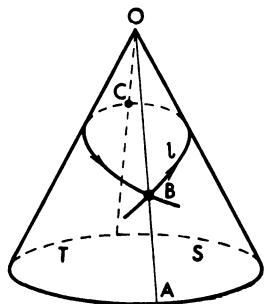


Fig. I.27

Consider point O' on a plane and straight line l' not passing through O' (Fig. I.28). If the plane is wrapped around the cone so that O' merges with vertex O of the cone, straight line l' will turn into geodesic line l .

Let C' be the base of a perpendicular dropped from O' onto l' . In wrapping the plane on the cone, ray $O'C'$ will turn into generatrix OC of the cone. Point C is sometimes called the *vertex of the geodesic line* on the conical surface. Denote the opposite generatrix of the cone by OA . Generatrices OA and OC divide the surface of the cone into two equal parts S and T . Cut the conical surface along generatrix OA and develop it on a plane so that vertex O again passes into point O' and generatrix OC into ray $O'C'$. This being done, geodesic line l will be again developed into straight line l' . The entire surface of the cone will turn into angle $A'O'A''$. Both its halves, S' and T' , will pass into halves S' and T' of the angle with straight line $O'C'$ as the bisector of this angle.

Let us consider two cases.

(1) Angle $A'O'A''$ (the developed cone angle) is greater than, or equal to, 180° (Fig. I.29). Straight line l' is entirely inside this angle. If the angle is again wrapped around the conical surface so that both sides of the angle $O'A'$ and $O'A''$ merge with generatrix OA , straight line l' will again

* Sometimes the double points are called *crunodes*.

turn into geodesic line l on the conical surface; different points of l' will pass into different points of the cone; hence l has no double points in this case.

(2) Angle $A'O'A''$ is less than 180° . Straight line l' perpendicular to bisector $O'C'$ intersects the sides of the angle at points denoted by B' and B'' (see Fig. I.28).

$\triangle B'O'B''$ is an isosceles triangle since its height $O'C'$ coincides with the bisector. Wrap angle $A'O'A''$ once again around the cone so that O' passes into the vertex of the cone,

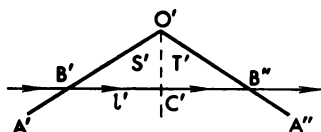


Fig. I.28

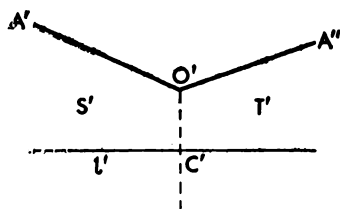


Fig. I.29

and both sides $O'A'$ and $O'A''$ of the angle into generatrix OA . Points B' and B'' owing to the equality of segments $O'B'$ and $O'B''$ will turn into one point B of this generatrix (see Fig. I.27). Straight line l' will pass into geodesic line l , segment $B'C'$ of l' lying in half S' of angle $B'O'B''$ will turn into arc \widehat{BC} of line l connecting points B and C and lying in half S of the conical surface. Similarly segment $B''C'$ lying in half T' of angle $B'O'B''$ will turn into arc \widehat{BC} of line l connecting points B and C and lying in half T of the conical surface. Point B is the double point of curve l .

Segment $B'B''$ of straight line l' will pass into arc \widehat{BCB} , which has the shape of a loop with coinciding ends.

Let us ascertain how many double points the geodesic has. The answer is furnished by the following theorem, which specifies the previous one.

Theorem 2. *Given a developed cone angle equal to α (α is the measure of the angle in degrees).*

(1) *If 180° is not divisible exactly by α , the number of double points of the geodesic is equal to a whole part of fraction $180/\alpha$.*

(2) *If 180° is divisible exactly by α , the number of double points is equal to $(180/\alpha) - 1$.*

If $\alpha > 180^\circ$, the whole part of fraction $180/\alpha = 0$; if $\alpha = 180^\circ$, then $(180/\alpha) - 1 = 0$. Hence, in accordance with

the theorem, in these cases the number of double points must be zero, which is a paraphrase of the first part of the previous theorem.

We must still consider the case when $\alpha < 180^\circ$. Using the notations of the previous theorem, angle $A'O'A''$ (Fig. I.30)

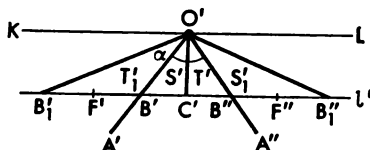


Fig. I.30

is the developed cone angle. Draw a perpendicular $O'C'$ to straight line l' through point O' , and straight line KL parallel to l' . Line KL divides the plane into two half-planes. Let us consider only that half-plane where straight line l' lies. Draw rays from point O' in this half-plane forming with ray $O'C'$ angles multiple to $\alpha/2$. These will be $O'B'$, $O'B''$, $O'B'_1$, $O'B''_1$, ... intersecting l' at points B' , B'' , B'_1 , B''_1 , ... Note that $O'B' = O'B''$, $O'B'_1 = O'B''_1$, ... Now roll the half-plane around the cone so that point O' coincides with vertex O of the cone, and ray $O'C'$ with generatrix OC (Fig. I.31). The angles of the half-plane equal to $\alpha/2$ and lying between neighbouring rays $O'B'_1$, $O'B'$, $O'C'$, $O'B''$, $O'B''_1$, ... will cover several times both halves S and T of the conical surface. Thus angle S' will pass into half S of the cone; the supplementary angles T'_1 and T'' into the other half T of the cone, and so on. Since ray $O'C'$ coincides with generatrix OC , rays $O'B'$ and $O'B''$ will coincide with the opposite generatrix OA , rays $O'B'_1$ and $O'B''_1$ again with OC , and so on.

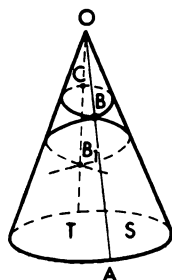


Fig. I.31

Since segments $O'B' = O'B''$, $O'B'_1 = O'B''_1$, the pairs of points B' and B'' , B'_1 and B''_1 , ... fall onto the same generatrix and will coincide pairwise: point B' will coincide with B'' and will get into point B of generatrix OA ; B'_1 and B''_1 will get into point B_1 of generatrix OC , and so on. Hence points B , B_1 , ... are double points of l into which straight

line l' has passed in wrapping the half-plane around the cone. The number of these points is equal to the number of rays $O'B'$, $O'B'_1$, \dots inside the right angle $KO'C'$. Since these rays form with $O'C'$ angles multiple to $\alpha/2$ and smaller than 90° , their number equals the number of quantities multiple to $\alpha/2$ and smaller than 90° (i.e. multiple to α and smaller than 180°). In other words, if 180° is not exactly divisible by α , the number of these rays is equal to a whole part of fraction $180/\alpha$. If, on the other hand, 180° is divisible by α , their number is equal to $(180/\alpha) - 1$.

To complete the proof of the theorem, it is necessary to show that all the double points of a geodesic are exactly those which are obtained through the merging of points B'_i and B''_i of straight line l' .

Indeed, a double point of geodesic l is obtained if two points of straight line l' pass into one and the same point of the cone when the half-plane is wrapped around the cone. For this both points must be equidistant from O' and lie on l' . It means that both points must lie on l' symmetrically with respect to C' . Let one of them, for example F' (see Fig. I.30), be on the left of C' and the other, F'' , on the right. If point F' does not coincide with either one of the points B' , B'' , B'_1 , B''_1 , \dots , it must lie inside one of the angles $C'O'B'$, $C'O'B''$, $B'O'B'_1$, $B'O'B''_1$, \dots denoted respectively by letters S'_i and T'_i in Fig. I.30. If point F' is inside angle T'_i , then point F'' symmetrical to it is inside angle S'_i , i.e. in wrapping the half-plane around the cone (see Fig. I.34), point F' will pass into a point lying inside semicone S , and point F'' into a point lying inside semicone T , and, vice versa, if point F' passes into a point lying inside semicone T , point F'' will pass into a point lying inside semicone S . In both cases F' and F'' will turn into different points of the cone. Hence, there are no other double points on geodesic l except those obtained by merging pairs B' and B'' , B'_1 and B''_1 , \dots . The theorem is proved.

Pay attention to the strip located between the parallel straight lines KL and l' . The reader is expected to find how this strip will be superposed on the conical surface with different magnitudes of the developed cone angle α (when $\alpha > 180^\circ$; $\alpha = 180^\circ$; $180^\circ > \alpha > 90^\circ$; $\alpha = 90^\circ$; $90^\circ > \alpha > 60^\circ$, etc.). In accordance with the reasoning at the end of the previous section, the stretched elastic thread lies on the conical surface along the geodesic line.

NOTE. Helixes, i.e. lines intersecting all the generatrices of a cone at one and the same angle α may also be considered (Fig. I.32). At $\alpha = 0$ and $\alpha = 90^\circ$ the helixes degenerate into generatrices and circular sections, respectively. At $\alpha \neq 0$ the helixes on the cone are not geodesics. In this they differ from helixes on a cylindrical surface.

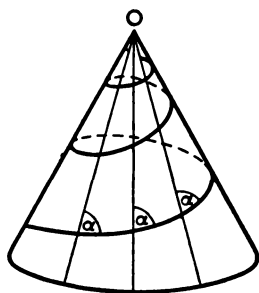


Fig. I.32

4. Clairaut's theorem for the geodesics on a cone. Let C be the top of geodesic s on the conical surface a distance $OC = c$ away from the vertex of the cone and a distance r_0 away from the axis of the cone (Fig. I.33). Then the geodesic at point C is normal to generatrix OC . Further, let A be an arbitrary point of the geodesic, r the distance from point A to the axis of the cone, α the angle between geodesic s and generatrix OA , and l the length of segment OA . The following ratio is obtained:

$$l \sin \alpha = c \quad (\text{I.1})$$

To prove formula (I.1) develop the conical surface on the plane (Fig. I.34). Generatrices OC and OA will pass into

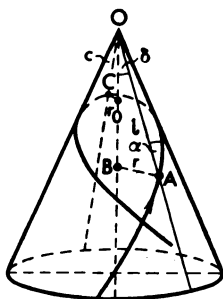


Fig. I.33

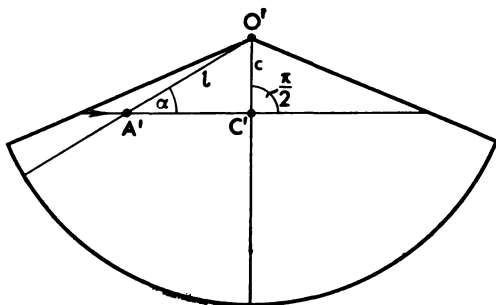


Fig. I.34

$O'C'$ and $O'A'$ (lengths c and l do not change), arc \widehat{AC} of geodesic s into segment $A'C'$ of the straight line ($O'C'$ is perpendicular to straight line $A'C'$), and the vertex angle A' in triangle $A'O'C'$ will be equal to α . From triangle $A'O'C'$

we obtain

$$l \sin \alpha = c$$

which is exactly what was to be proved.

Note that if δ is the angle between the generatrix and the axis of the cone (see Fig. I.33), then $r = l \sin \delta$. Multiplying both parts of ratio (I.1) by $\sin \delta$, we obtain

$$l \sin \delta \times \sin \alpha = c \sin \delta$$

or

$$r \sin \alpha = c_1 \quad (\text{I.2})$$

where $c_1 = c \sin \delta$ is a constant for the geodesic.

The latter equation can be proved by the following

Theorem 3. *For all points A of geodesic s on the conical surface the expression $r \sin \alpha$, where r is the distance between point A and the axis of the cone and α is the angle between generatrix OA and geodesic s, is constant:*

$$r \sin \alpha = \text{const} \quad (\text{I.3})$$

This theorem is a particular case of Clairaut's theorem (see Sec. III.3).

A cylinder may be considered as the limiting case of a cone (when the vertex of the cone tends to infinity). The helix on a cylinder is the counterpart of the geodesic line on a cone. Formula (I.3) obviously holds true for the cylinder as well: distance r (from the axis) is identical for all the points of the cylinder, angle α between the helix and the generatrices of the cylinder is also the same for all the points of the helix.

I.4. Shortest Lines on a Spherical Surface

1. The length of a line. In examining the shortest lines on the surface of a cylinder or a cone we took advantage of the fact that cylindrical and conical surfaces can be developed onto a plane. This method won't work in studying the shortest lines on the surface of a sphere, which cannot be rolled out onto a plane.

The shortest of all the lines with fixed ends is found in elementary geometry from a theorem which states that a side of a triangle is less than the sum of the other two sides. On the basis of this theorem it can be proved that segment AB of a straight line is shorter than any broken line $A_0 A_1 A_2 \dots A_{n-1} A_n$ having the same ends $A_0 = A$ and $A_n = B$ (Fig. I.35). Indeed, we are able only to shorten the broken

line if we replace its two adjacent segments A_0A_1 and A_1A_2 with segment A_0A_2 (since base A_0A_2 of triangle $A_0A_1A_2$ is shorter than the sum of sides A_0A_1 and A_1A_2).^{*} In doing so we replace broken line $A_0A_1A_2 \dots A_{n-1}A_n$ with broken line $A_0A_2 \dots A_{n-1}A_n$ having one side less. Similarly, in this broken line two adjacent segments A_0A_2 and A_2A_3 can be replaced by one segment A_0A_3 which will not increase

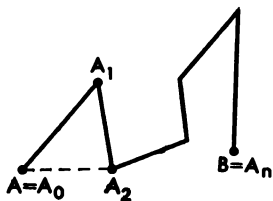


Fig. I.35



Fig. I.36

the length of the broken line. Thus we obtain broken line $A_0A_3 \dots A_{n-1}A_n$ which has one side less. Thus we can successively decrease the number of segments of the broken line until it is reduced to only one, namely, segment $A_0A_n = AB$. In passing from one broken line to another its length could only decrease (sometimes the length remained unchanged, though it could not remain unchanged in each of the transitions since this is possible only if all the points A_0, A_1, \dots, A_n lie on one straight line AB , which is ruled out in this case). It follows that the original broken line was longer than segment AB . In elementary geometry it is merely proved that segment AB of a straight line is shorter than any broken line connecting the same points A and B .

To draw a similar conclusion for an arbitrary line connecting points A and B , first of all the length of the curve must be found. In elementary geometry the length of a circle is found as the limit value of the lengths of inscribed polygons when the number of sides of the polygon tends to infinity and the length of the largest side tends to zero.

The length of an arbitrary line can be found in a similar way. Consider a line q connecting points A and B (Fig. I.36).

^{*} If points A_0, A_1 and A_2 lie on the same straight line, the sum of the lengths of two segments A_0A_1 and A_1A_2 is equal to the length of segment A_0A_2 . Therefore by replacing the two segments A_0A_1 and A_1A_2 with one segment A_0A_2 we do not increase the length of the broken line. This remark is also used in further discussion.

Let us move along this line from A to B and successively mark $(n + 1)$ points: $A_0 = A, A_1, A_2, \dots, A_n = B$. Connect these points one after another by segments. We obtain broken line $A_0A_1A_2 \dots A_n$, which we shall call a *broken line inscribed into a curve*. Now let us construct broken lines inscribed into curve q , i.e. lines with an *infinitely growing number of sides*, so that the *length of the largest side tends to zero*. It can be shown that under such conditions *the lengths of the inscribed polygons tend to a limit that is assumed to be the length of the line*.

Since segment AB is shorter than any broken line connecting points A and B , and the length of a curve connecting these points is the limit of the lengths of the broken lines, it follows that the straight-line segment is the shortest line among all curves connecting A and B .

2. **The shortest line on a spherical surface.** Proceed now to find the shortest lines on a spherical surface. Note that only one great circle can be drawn through two points A and B on the spherical surface, if they do not lie on the opposite endpoints of a diameter. An infinite number of curves can be drawn through two points lying on the opposite endpoints of one and the same diameter. For a while we agree to exclude the latter case without any provisos, in short, in discussing two points on a spherical surface we shall tacitly presume that these two points do not lie on one diameter of the sphere.

Draw a great circle passing through two given points A and B on a spherical surface. Points A and B (since they do not lie on the ends of a diameter) divide the great circle into two unequal arcs. Denote the minor of the arcs by \widehat{AB} .

Consider three points on the spherical surface, viz. A, B and C , connected by the arcs of great circles, $\widehat{AB}, \widehat{BC}$ and \widehat{CA} . These three arcs form a so-called *spherical triangle* ABC in which arcs $\widehat{AB}, \widehat{BC}$ and \widehat{CA} are its *sides*.

It turns out that the basic theorem concerning the lengths of the sides of a plane triangle holds for spherical triangles as well.

Theorem. *Each side of a spherical triangle is shorter than the sum of the other two sides.*

Consider a spherical triangle ABC on the surface of a sphere with the centre at point O (Fig. I.37). Side \widehat{AB} of

this triangle is an arc of a great circle, i.e. a circle with the centre at point O . In the plane of this circle central angle AOB corresponds to arc \widehat{AB} . Similarly, in the planes with sides \widehat{BC} and \widehat{CA} the corresponding central angles are BOC

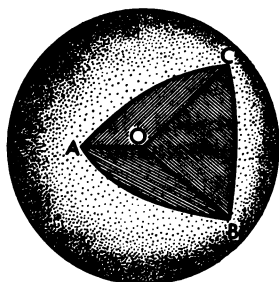


Fig. I.37

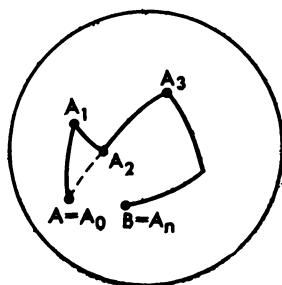


Fig. I.38

and COA . The lengths of sides \widehat{AB} , \widehat{BC} and \widehat{CA} , as arcs of great circles of equal radii, are proportional to central angles AOB , BOC and COA .

The three planes of the great circles form a trihedral angle with the vertex at point O and with plane angles AOB , BOC and COA . The lengths of the sides of the spherical triangle are proportional to the respective plane angles of the trihedral angle. And since in a trihedral angle each plane angle is less than the sum of the other two plane angles, the similar inequality will also be valid for the sides of the spherical triangle, the sides being proportional to the plane angles. This proves the theorem.

Consider a sequence of points $A_0, A_1, A_2, A_3, \dots, A_n$ on a spherical surface; the points are connected by the arcs of great circles: $A_0A_1, A_1A_2, A_2A_3, \dots, A_{n-1}A_n$. A set of these arcs is called a *spherical broken line* connecting points A_0 and A_n (Fig. I.38).

From the premise that a side of a plane triangle is less than the sum of the other two sides there followed the theorem that segment AB of a straight line is shorter than a broken line connecting the same points A and B . Similarly, from the premise that one side of a spherical triangle is less than the sum of the other two sides it follows that arc \widehat{AB} of a great circle is shorter than any broken line connecting the same points. Further, for a spherical surface just as for

a plane, the lengths of curves connecting points A and B are the limiting lengths of spherical broken lines connecting these points. Since arc \widehat{AB} of the great circle is shorter than all the spherical broken lines connecting A and B , it is shorter than all the curves connecting these points.

The proof that arc \widehat{AB} is shorter than any other broken line connecting points A and B repeats in the main the proof of a similar theorem for a broken line lying on a plane.

Consider arc \widehat{AB} and broken line $A_0A_1A_2A_3 \dots A_n$, where $A_0 = A$ and $A_n = B$. In spherical triangle $A_0A_1A_2$ side $\widehat{A_0A_2}$ is less than the sum of two sides $\widehat{A_0A_1}$ and $\widehat{A_1A_2}$.^{*} Replace two sides $\widehat{A_0A_1}$ and $\widehat{A_1A_2}$ with arc $\widehat{A_0A_2}$. We have a new line $A_0A_2A_3 \dots A_n$ which is perhaps shorter than the original one and contains one segment less. Next, replace two sides $\widehat{A_0A_2}$ and $\widehat{A_2A_3}$ with one side $\widehat{A_0A_3}$. This will either shorten the length of the broken line or leave it unchanged. Similar transformations (the replacement of two adjacent segments of the broken line with one) will be further continued. Each reduction in the number of sides will either reduce the length of the broken line or leave it unchanged. Then each time we shall have new broken lines connecting A and B with a fewer number of segments, and finally we shall come to a broken line consisting of only one segment, i.e. of arc \widehat{AB} . Each time the length of the broken line either decreases or remains unchanged. But the length of the broken line cannot remain unchanged in each step, since this would mean that points A_0, A_1, \dots, A_n lie on one great circle on arc \widehat{AB} , which is ruled out in this case. Therefore, *the length of the initial broken line $A_0A_1 \dots A_n$ is greater than the length of \widehat{AB} .*

Now consider a case when points A and B are at the end-points of one and the same diameter of a sphere. In this case there is an infinite number of arcs of great circles connecting

^{*} If points A_0, A_1 and A_2 lie on a great circle, side $\widehat{A_0A_2}$ is either equal to the sum of sides $\widehat{A_0A_1}$ and $\widehat{A_1A_2}$ if this sum is less than a semicircle or is smaller than it if the sum is greater than the semicircle. Therefore, in replacing two sides $\widehat{A_0A_1}$ and $\widehat{A_1A_2}$ with one $\widehat{A_0A_2}$, the length of the broken line may either become less or remain unchanged. This note also concerns further discussion.

A and B and having AB as a diameter. All of them are of equal length. On the other hand, any other curve q connecting the same points A and B is of a length greater than that of a semicircle of the great circle. Indeed, let point C (different from A and B) lie on q and divide this line into two lines (AC) and (CB) . Draw a semicircle of the great circle \widehat{ACB} ; it consists of two arcs \widehat{AC} and \widehat{CB} . Each of these arcs is shorter than any other curve on the spherical surface connecting the same points A and B . Since our curve q is not a semicircle, at least one of its parts (AC) or (CB) does not coincide with the respective arc \widehat{AC} or \widehat{CB} . For example, let (AC) not coincide with \widehat{AC} . Then the length of (AC) is greater than that of \widehat{AC} . Further, the length of (CB) is either greater than that of \widehat{CB} (if they do not coincide) or is equal to it [if (CB) coincides with \widehat{CB}]. It follows that the total length of q is greater than the length of \widehat{ACB} .

For two diametrically opposite points A and B there exists an infinite number of shortest curves connecting these points; all the semicircles of the great circles connecting A and B are just these curves.

3. An additional observation. A spherical surface cannot be developed on part of a plane without deformation, i.e. without changing the length of lines lying on it. However, it is possible to develop onto a plane a very narrow strip stretched on the spherical surface along some line q with only an infinitesimal deformation of the length of the lines lying on this strip. The narrower the strip on the sphere the smaller are the distortions and the greater the accuracy of developing the strip onto a plane. Using the language of the theory of limits, the deformation of the length of lines on the strip is a magnitude of higher order of smallness in comparison with the width of the strip.

If a narrow strip situated on a spherical surface is developed onto a plane, an arc of a great circle contained in the strip passes into a straight-line segment (and vice versa).

Indeed, arc \widehat{AB} of the great circle on a spherical strip is the shortest of all the other arcs lying on the strip and connecting A and B . If in developing the strip onto a plane points A and B pass into A' and B' , arc \widehat{AB} will turn into an arc connecting A' and B' on the plane, the arc being shorter than the neighbouring plane arcs connecting the same points. Hence, \widehat{AB} will pass into segment $A'B'$.

COROLLARY. On a spherical surface cut a narrow strip around a great circle and develop the strip onto a plane. This strip will turn

into a flat straight strip and the great circle into the midline of the strip. Vice versa, if a narrow flat strip (tape) is wrapped onto a spherical surface, it will lie on this surface along the great circle (Fig. I.39).

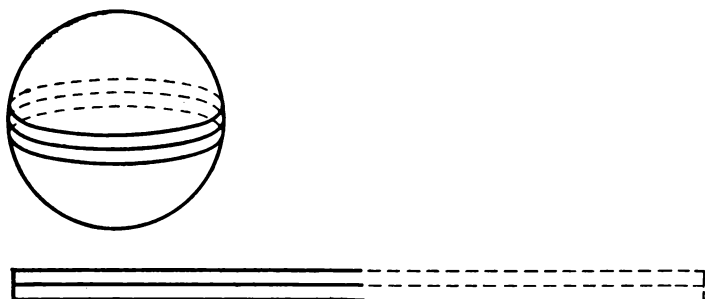


Fig. I.39

Let us see what a narrow strip containing the arc of a small circle q (i.e. a circumference on a spherical surface other than the great circle) will turn into.

First note the following. Cut a conical surface by a plane normal to the axis of a cone. This plane will cut the conical surface along cir-

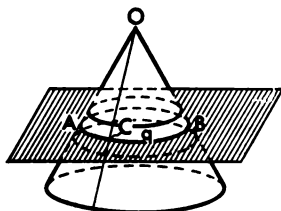


Fig. I.40

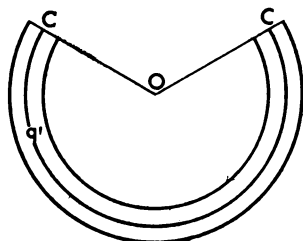


Fig. I.41

cumference q . All the segments of the generatrices passing from vertex O of the cone to circumference q are equal (for example, in Fig. I.40 $OA = OB = OC$). If the conical surface is cut along generatrix OC and this surface is developed onto the plane, circumference q will pass into arc q' of a radius equal to OC . The narrow strip on the surface of the cone, the one with circumference q as its midline, will be developed onto a plane and become a strip with arc q' as the midline (Fig. I.41).

Again consider a spherical surface (Fig. I.42). Choose diameter AB passing through centre O_1 of small circle p_1 and centre O of the sphere, then draw great circle p of diameter AB crossing small circle p_1 at point C . Let r be the radius of p_1 , R the radius of the sphere and α the angle O_1CO . Then

$$\cos \alpha = \frac{r}{R}$$

Draw a tangent CD to p at point C until it intersects at point D the extension of diameter AB . We have $\angle CDO = \angle O_1CO = \alpha$ (owing to the perpendicularity of the sides of these angles). From triangle OCD we have:

$$CD = R \cot \alpha = R \frac{\cos \alpha}{\sqrt{1 - \cos^2 \alpha}} = \\ = R \left[\frac{r}{R} : \sqrt{1 - \left(\frac{r}{R}\right)^2} \right] = \frac{rR}{\sqrt{R^2 - r^2}}$$

Rotate the drawing about axis AB . Straight line CD will generate a conical surface, and circumference p will circumscribe a sphere of radius R . The conical and spherical surfaces will touch each other along circumference p_1 .

Minor arc $\widehat{C_1C_2}$ of circle p , i.e. the arc containing point C , may be considered congruent with the minor segment of the tangent.* If the arc is rotated about AB , it will circumscribe a spherical strip containing small circle p_1 . This strip may be considered congruent with the strip on the cone**, which touches the sphere along circumference p_1 (this strip on the conical surface is formed by the rotation of a segment of the tangent which is considered congruent with arc $\widehat{C_1C_2}$). If this strip is cut along $\widehat{C_1C_2}$ and is developed onto a plane, circumference p_1 will pass into an arc of a circumference of a radius equal to CD , i.e. of a radius

$$l = \frac{Rr}{\sqrt{R^2 - r^2}}$$

and the narrow strip of the spherical surface, i.e. the strip having circumference p_1 as its midline, will be developed into a flat strip enclosing the arc of the circumference of radius l .

Vice versa, if a narrow flat strip with the arc of the circumference of radius l as its midline is wrapped around the spherical surface of radius R , it will lie on the spherical surface along the small circle. The radius of the circle can be found from the equation

$$l = \frac{Rr}{\sqrt{R^2 - r^2}}$$

Whence it is easy to see that

$$r = \frac{Rl}{\sqrt{R^2 + l^2}}$$

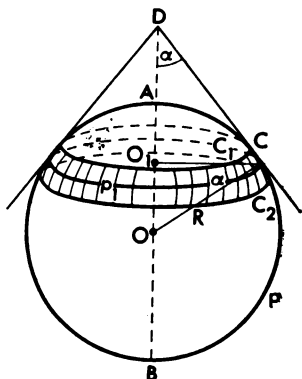


Fig. 1.42

* Congruent if the values of the highest order of smallness are neglected in comparison with the length of $\widehat{C_1C_2}$.

** Congruent in the same sense.

CHAPTER II

Some Properties of Plane and Space Curves and Associated Problems

II.1. Tangent and Normals to Plane Curves and Associated Problems

1. A tangent to a curve. Given curve q on a plane or in space and point A on it (Fig. II.1). Consider another point B on the same curve. Connect points A and B by straight line n . The straight line is called a *secant*. Bring point B nearer to point A moving it along curve q . In the process, secant n

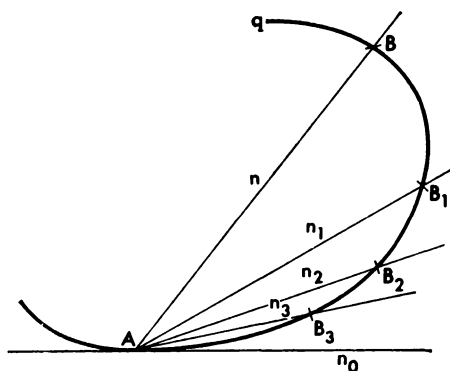


Fig. II.1

will rotate about point A . When point B occupies points B_1, B_2, B_3, \dots , secant n will coincide with straight lines AB_1, AB_2, AB_3, \dots . If point B tends to point A , secant n tends to its limit position, viz. to straight line n_0 . This limit position of a secant is called the *tangent* to curve q at point A .

Imagine a mass point moving along curve q and suddenly breaking loose at point A . After this it will move by inertia along tangent n_0 to the curve at point A .

2. **A normal.** Now suppose that curve q lies in a plane (we shall call such a curve a *plane curve*). Straight line MN , passing through point A and perpendicular to tangent n_0 to curve q at this point, is called a *normal* to curve q at point A (Fig. II.2).

3. **The shortest distance between two curves.** Consider point A that can move only along curve q . Let P be the resultant of forces acting on point A (Fig. II.3). Resolve force P into two components: the *tangential component* P_1 directed along the tangent to curve q at point A and the *normal component* P_2

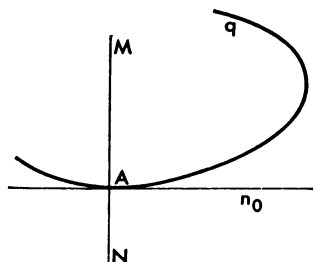


Fig. II.2

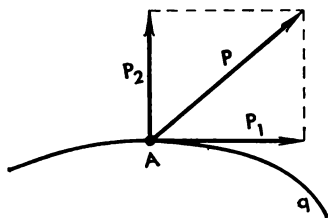


Fig. II.3

directed along the normal to curve q at point A . The tangential component moves point A along curve q . For this reason *point A is in equilibrium if the tangential component P_1 is absent, i.e. if P coincides with P_2 . Hence, force P acts along the normal to curve q at point A .*

Consider two curves q and q_1 . We must find the shortest of lines r whose one end, A , is on curve q and the other, B , on curve q_1 (Fig. II.4). Assume lines q and q_1 to be immovable and rigid. Consider elastic thread r whose one end, A , slides along curve q and the other, B , along curve q_1 (we can imagine that at point A there is a small ring through which curve q is threaded and another ring at point B with q_1 threaded through it, and that the ends of the thread are attached to these rings). Thread r tends to occupy a position in which its length is the least. Let A_0B_0 be that position. The thread is then in equilibrium. Obviously, A_0B_0 is a straight-line segment connecting points A_0 on q and B_0 on q_1 (if this line were not a straight-line segment, it could be made shorter with the position of its ends not changed). Since the line in position A_0B_0 is in equilibrium, its end A_0 is also in equilibrium. Point A_0 is under a tensile strength directed along

segment A_0B_0 . Owing to the equilibrium condition of a point on a curve (derived previously), segment A_0B_0 is a normal to curve q at point A_0 . Similarly, it can be shown that this segment is also a normal to curve q_1 at point B_0 .

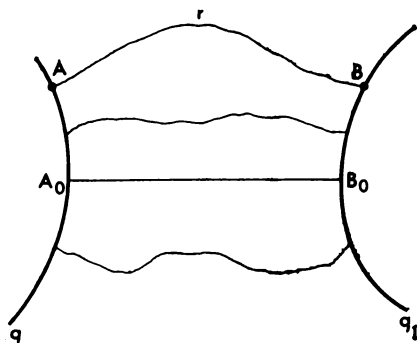


Fig. II.4

Thus, *the shortest of the lines connecting points of two curves is the common normal to these curves.*

Similarly, *the shortest of the lines connecting point A with curve q is the normal to curve q drawn from point A.*

4. The reflection problem. Let q be a fixed curve. Consider all kinds of curves ACB connecting two given points

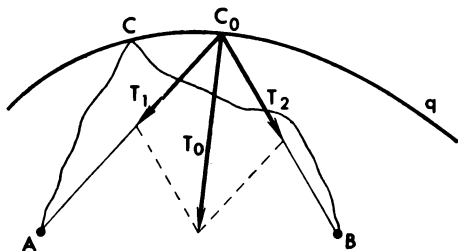


Fig. II.5

A and B and having a common point C with curve q , or, so to say, the curves connect points A and B reflecting at curve q .

Consider thread \widehat{ACB} fixed at ends A and B with point C moving along curve q (Fig. II.5).

Let AC_0B be the shortest of the lines connecting points A and B and reflecting at curve q (C_0 is a point on curve q). The thread in position AC_0B is in equilibrium.

Obviously, both parts AC_0 and C_0B of the shortest curve are straight-line segments. Point C_0 of the thread on the curve is in equilibrium because this point is under two tensile strengths equal in magnitude*: force T_1 is directed along segment C_0A , and force T_2 along segment C_0B . Their resultant T_0 acts along the bisector of angle AC_0B . Owing to the equilibrium condition, T_0 is directed along the normal to curve q at point C_0 . Hence, the bisector of angle AC_0B is the normal to curve q at point C_0 .

The shortest of the curves connecting points A and B and reflecting at curve q , is broken line AC_0B with vertex C_0 on curve q at which the normal to this curve coincides with the bisector of angle AC_0B .

5. The shortest distances in a region. Consider a region bounded by some lines on a plane. The region may be finite

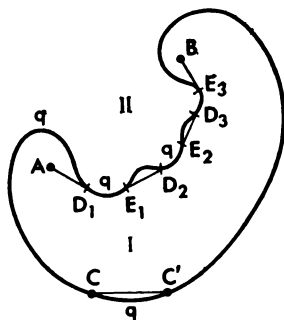


Fig. II.6

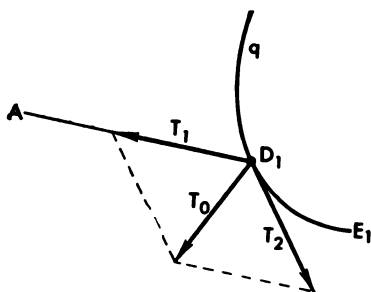


Fig. II.7

(region I in Fig. II.6) or infinite (for example, region II obtained by eliminating region I from the plane).

Let us find the shortest of the lines connecting points A and B in region I . This line \widehat{AB} is the position of equilibrium of a flexible thread lying in the region and fixed at points A and B (we shall regard the boundary of the region as enclosed). The thread may contain parts of boundary q of region I .

* Tensile strength is the same at all points of the thread.

Let $s_0 = AD_1E_1D_2E_2 \dots D_nE_nB$ be the shortest of lines s . It consists of parts $\widehat{E_1D_1}$, $\widehat{E_2D_2}$, \dots , $\widehat{E_nD_n}$ of the boundary (in Fig. II.6 $n = 3$) and of lines AD_1 , E_1D_2 , \dots , E_nB lying entirely (except the ends) inside I . Obviously, lines AD_1 , E_1D_2 , \dots , E_nB are straight-line segments.

Each part D_1E_1 , D_2E_2 , \dots , D_nE_n of the boundary entering s_0 bulges towards I . Indeed, for each sufficiently small area $\widehat{CC'}$ of boundary q bulging towards II , chord CC' lies in I . This chord is shorter than arc $\widehat{CC'}$. Therefore, if line s_0 did contain the arc $\widehat{CC'}$ of the boundary, we would be able to shorten s_0 replacing arc $\widehat{CC'}$ with chord CC' lying in I .

Hence, the shortest line can contain only the parts of the boundary bulging towards I .

Segments AD_1 , E_1D_2 , \dots , $E_{n-1}D_n$, E_nB included in s_0 touch curve q at points D_1 , E_1 , D_2 , E_2 , \dots , D_n , E_n (see Fig. II.6), respectively.

Indeed, two parts of the thread, namely segment AD_1 and part $\widehat{D_1E_1}$ of curve q , converge at a point, D_1 for example. Tension T_1 of part AD_1 is directed along segment D_1A (Fig. II.7) and tension T_2 of part $\widehat{D_1E_1}$ along the tangent to q at point D_1 . If the angle between the directions of T_1 and T_2 is other than 180° , the resultant T_0 of forces T_1 and T_2 will displace point D_1 (see Fig. II.7), i.e. the thread will not be in equilibrium. This angle is equal to 180° , i.e. segment AD_1 touches q at point D_1 .

Thus the shortest line in region I , i.e. the line connecting points A and B , consists of segments of tangents AD_1 , E_1D_2 , \dots , E_nB and parts of boundary D_1E_1 , D_2E_2 , \dots , D_nE_n bulging towards I .

In considering the shortest lines on a polyhedral surface (see the footnote (*) on p. 12) a proviso was made concerning the location of a straight line on the development. On the basis of the material presented in this section this restriction may be dropped.

II.2. Some Information on the Theory of Plane and Space Curves

1. Osculating circles. Let q be a plane curve (Fig. II.8). Draw tangent KL and normal MN at point A of this curve. Also draw different circles touching straight line KL at

point A (i.e. having a common tangent with curve q at point A). It is obvious that their centres lie on normal MN .

Among all these circles one is the nearest to curve q at point A . In our drawing it is circle r . This circle is called an *osculating circle*. To a certain degree, minor arc \widehat{BC} of curve q containing point A may be considered the arc of osculating circle r . The smaller the arc \widehat{BC} the greater is the accuracy

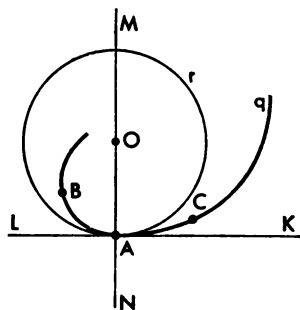


Fig. II.8

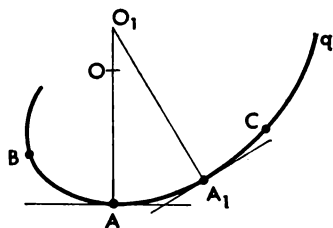


Fig. II.9

with which it can be replaced by the arc of circle r . Point O , being the centre of circle r , is sometimes termed the *centre of curvature*. Thus minor arc \widehat{BC} of curve q containing point A may, to a certain approximation, be called the arc of a circle with the centre in the centre of curvature, i.e. point O .

The centre of a circle lies on the intersection of its two radii, and since radii are normals of a circle, we may say that the centre of a circle lies on the intersection of its normals.

Consider an arbitrary curve q with point A on it and minor arc \widehat{BC} including this point (Fig. II.9). This arc may be approximately considered an arc of an osculating circle at point O . The task is to find the centre of this circle.

Since arc \widehat{BC} can be roughly considered as that of an osculating circle, the following technique of constructing the centre of curvature may be recommended. Draw normals to curve q at point A and at some point A_1 of the curve close to it. These normals will intersect at point O_1 . If we consider arc \widehat{BC} as that of an osculating circle, point O_1 in accordance with the preceding reasoning will be the centre of the osculating circle (the centre of curvature).

NOTE. Our construction of the centre of the osculating circle will be an approximate one. The smaller the arc \widehat{BC} the more accurate will be the construction. It is possible to determine the centre of curvature of curve q at point A accurately as the limiting position to which the point of intersection of the normal at point A with the normal at the neighbouring point A_1 tends when point A_1 approaches point A . The nearer point A_1 is to point A the closer is the point of intersection of the two normals, O_1 , to the limiting position, i.e. point O . The osculating circle may be defined as the circle of radius OA with the centre at O .

Example. The above approximate method is employed in Fig. II.10 for constructing the centres of curvature and osculating circles at points B and A of an ellipse.

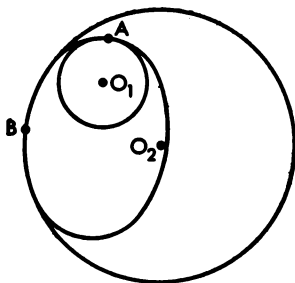


Fig. II.10

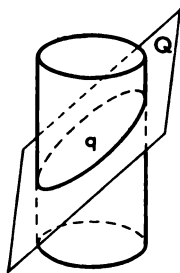


Fig. II.11

2. Space curves. Up till now we have discussed plane curves. In this section we shall consider space curves. Let us draw attention to the fact that certain curves, such as helixes, cannot lie in a plane.

Let q be a helix on a cylindrical surface. If q were in plane Q , it would be the line of intersection of this plane with the cylinder. Two cases are possible: either plane Q intersects the axis of the cylinder or it is parallel to the axis. If the plane intersects the axis of the cylinder, it will cut the cylinder along a closed curve (along an ellipse as in Fig. II.11) and not along a helix, which is an open curve. If, on the other hand, the plane is parallel to the axis of the cylinder, it will cut its surface along two straight lines or, touching the cylindrical surface, will have one common straight line with it or, finally, it will not intersect the cylinder at all. In any case a helix cannot be the line of intersection of a plane with the cylindrical surface.

A *tangent* to a space curve is determined in the same way as to a plane curve. Let us call any straight line passing through point A and perpendicular to the tangent at point A

a *normal* to a space curve q at point A . But an infinite number of perpendiculars can be dropped onto a straight line at any of its points. Therefore, there is an infinite number of normals to curve q at point A . They fill in a whole plane perpendicular to the tangent at point A (Fig. II.12).

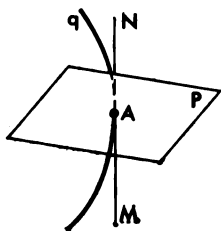


Fig. II.12

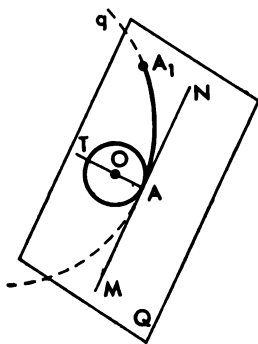


Fig. II.13

3. Osculating planes. Consider point A on curve q and straight line MN tangent to curve q at this point (Fig. II.13). Let A_1 be a point on the curve very close to point A . Minor segment $\widehat{AA_1}$ of the space curve q can be approximately considered as the arc of a plane curve. Plane Q passing through tangent MN and through point A_1 can be approximately considered as a plane in which minor arc $\widehat{AA_1}$ of the curve lies. Plane Q is called the *osculating plane* to curve q at point A .

NOTE. Let us give an accurate definition of an osculating plane. Draw plane Q' passing through tangent MN to the curve at point A and through another point A_1 of the same curve. If point A_1 tends to point A moving along curve q , plane Q' will turn about MN and tend to a limiting plane Q , which is called the *osculating plane*. If point A_1 is very close to point A , plane Q' , passing through MN and point A_1 , will be very close to Q . For this reason we may roughly consider plane Q' to be the osculating plane.

4. The principal normal. The *principal normal* to curve q at point A is normal AT lying in the osculating plane (see Fig. II.13).

If curve q is entirely in plane Q (i.e. if q is a plane curve), plane Q is the osculating plane for all the points of curve q ,

and normals to q lying in this plane are its principal normals.

5. Osculating circles for space curves. A minor arc of a space curve with point A can be roughly considered as a plane arc which lies in plane Q osculating with curve q at point A . But each plane arc, in turn, can be roughly considered as the arc of an osculating circle (situated in the same plane and having a common tangent with the curve). It means that the minor arc of curve q with point A can be approximately considered as the arc of some circle in the osculating plane (see Fig. II.13). This circle is called the *osculating circle of the space curve*. Its centre O lies on the principal normal to the curve. Thus, minor segments of plane and space curves can be roughly regarded as the arcs of osculating circles. *The smaller the arc of a curve the greater will be the accuracy with which the arcs of the curve can be replaced by the arcs of the osculating circles.*

The above information from the theory of curves will be required for further discussion.

II.3. Some Information on the Theory of Surfaces

1. A tangent plane and a normal to a surface. Consider surface S and point A on it (Fig. II.14). A minor segment of

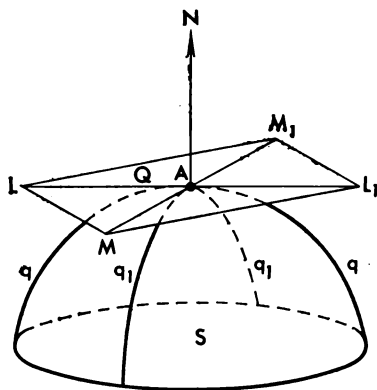


Fig. II.14

the surface around point A can be approximately regarded as a segment of plane Q , the so-called tangent plane to sur-

face S at point A . *Tangent plane* Q is a plane which contains tangents at point A to the curves lying on surface S and passing through point A .

If two curves q and q_1 are drawn on S so that they pass through point A with non-osculating tangents LL_1 and MM_1 drawn at the point, then tangent plane Q is a plane defined by straight lines LL_1 and MM_1 .

A *normal* to surface S at point A is a straight line passing through A and perpendicular to tangent plane Q at point A of surface S .

Normal AN to the surface is the normal to all the curves lying on this surface and passing through point A (generally speaking, it will not be their principal normal at this point).

Examples. The normal to a spherical surface at some point of the surface is the radius of the sphere at this point.

The normal to a cylindrical surface at some point of the surface is the radius of the circular section of the cylinder at this point.

NOTE. A curve may not necessarily have a tangent at each of its points. Take for instance a broken line. It is impossible to draw a tangent to its vertex. Similarly, a space curve may not have an osculating plane, and a surface—a tangent plane and normal, etc. For example, a conical surface has no tangent plane and normal to the vertex of the cone.

Henceforth we shall restrict ourselves to *smooth curves*, i.e. curves with a tangent at each of its points as well as an osculating plane and a centre of curvature, and to *smooth surfaces*, i.e. surfaces which may have a normal to each of its points. On surfaces only smooth curves will be considered.

2. Equilibrium condition for a point on a surface. Consider point A capable of moving only along surface S . Let P represent the resultant of forces acting on this point (Fig. II.15). Denote by P_1 the tangential component of force P (i.e. the component lying in plane Q tangent to S at point A) and by P_2 the normal component (i.e. the component drawn along the normal to surface S at point A).

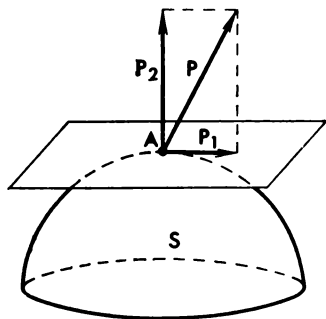


Fig. II.15

The tangential component P_1 moves point A along the surface; that is why for point A on the surface to be in equilibrium the tangential component P_1 must be zero. It means that force P coincides with its normal component P_2 . So *for point A on a surface to be in equilibrium the resultant P of all the forces acting on point A must be directed along the normal to the surface at this point.*

3. Some problems connected with shortest lines in space. Find the shortest line connecting the points of two space curves.

From what has been said in Item 3 of Sec. II.1 one can see that the shortest line connecting points of two curves is a segment of their common normal.

In particular, the line representing the shortest distance between points of two non-intersecting lines in space is a segment of their common perpendicular.

Similarly, it can be proved that the shortest distance between two surfaces is a segment of their common normal.

CHAPTER III

Geodesic Lines

III.1. Bernoulli's Theorem on Geodesic Lines

1. Equilibrium of an elastic thread on a surface. Given two points A and B on surface S . These points can be connected by an infinite number of lines lying on the surface. One of them is the shortest line q . Our task is to examine the properties of this line.

Imagine a rubber thread stretched on the surface and fixed at points A and B (Fig. III.1). This thread is in equilibrium if it assumes the shape of the shortest line q . Indeed, if we

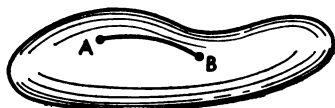


Fig. III.1

disturb the equilibrium position q by somewhat changing the thread's shape, we shall lengthen it and, striving to shorten, it will again assume position q . Consequently, *a thread lying along the shortest line q will be in a state of equilibrium, a stable equilibrium at that.*

We shall start with examining the line of equilibrium of an elastic thread on a surface.

First consider thread \widehat{AB} having the shape of an arc of a circle (Fig. III.2). Section \widehat{CD} of this thread is under the tension acting from the other sections of the thread, namely, point C is under the effect of tensioning of thread section CA and point D of thread section DB . These tensions act along tangents at points C and D . Denote them by P_1 and P_2 .

Forces P_1 and P_2 are equal in magnitude, otherwise part \widehat{CD} of the thread would not be in equilibrium. Find the resultant of forces P_1 and P_2 .

Let point M be the point of intersection of the tangents at points C and D (forces P_1 and P_2 act along these tangents). Transfer forces P_1 and P_2 to point M . It is obvious that the resultant acts in the direction of centre O of the circle (with thread

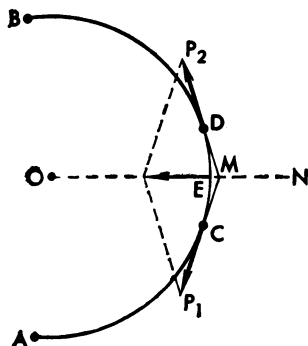


Fig. III.2

\widehat{AB} lying on it). Denote the midpoint of arc \widehat{CD} by E . The resultant of the tensions acting on arc \widehat{CD} passes through the midpoint E of the arc and is directed along radius EO . Since radius EO is the normal to arc \widehat{AB} at point E , a final conclusion can be drawn, i.e. *the resultant of the tensions acting on arc \widehat{CD} of the circle passes through midpoint E of the arc and is directed*

along the normal to the circle at point E .

Consider now a general case. A rubber thread stretched on a surface is fixed at points A and B and has the shape of curve q .

Isolate minor section \widehat{CD} of this thread.* Section \widehat{CD} is under the action of tensions P_1 and P_2 applied at points C and D in the direction of the tangents to q at these points. The minor arc of the curve can be considered as the arc of an osculating circle at midpoint E of the arc. Radius EO of the circle is directed *along the principal normal to curve q at point E* . The resultant of the tensions acting on the arc of the circle will be directed along the radius passing through the midpoint of the arc; in our case, along radius EO . So *the resultant of the tensions acting on minor arc \widehat{CD} of the thread passes through the midpoint E of the arc and is directed along principal normal EO at point E .*

* In view of its smallness \widehat{CD} can be considered as the arc of a circle, and Fig. III.2 may be used for this case.

Now it is not difficult to find the conditions for the thread to be in equilibrium. If the thread is in a state of equilibrium, each of its minor sections \widehat{CD} is also in a state of equilibrium. For arc \widehat{CD} to be in equilibrium, the resultant must be directed along the normal to the surface. The tensions acting on \widehat{CD} have a resultant directed along principal normal EO to curve q . It means that one and the same straight line EO must be the principal normal to curve q at point E and also the normal to surface S at this point.

The following theorem ensues: *For elastic thread q stretched on surface S to be in a state of equilibrium, it is necessary that at any of its points A the principal normal to it coincides with the normal to the surface.*

2. Geodesic lines. *Line q is called the geodesic on surface S if at each of its points the principal normal coincides with the normal to the surface.*

The geodesic line can be defined as a line on a surface which at each of its points has an osculating plane passing through the normal to the surface at this point. Indeed, let A be a point on curve q lying on surface S . The normal to the surface at point E is at the same time the normal to curve q at this point; this normal is the principal one if it lies in the plane osculating with q at point A .

The theorem can be formulated as follows: *a stretched thread on a surface will be in the state of equilibrium if it lies along the geodesic line of this surface.*

Example 1. As has been proved, stretched threads on the surface of a cylinder lie along the helixes. Therefore, *helixes are the geodesic lines on the surface of a cylinder.* The principal normals to the helixes coincide with the normals to the cylindrical surface, and the normals to the cylindrical surface are the radii of circular sections. So *the principal normals of helixes are the radii of circular sections.*

Example 2. Let us see in what case plane curve q can be the geodesic line on some surface S . Denote by Q the plane with line q . For plane curve q the osculating plane at any of its points will be plane Q itself.

Owing to the second definition of the geodesic line, if curve q is the geodesic line, the normals to surface S at points of curve q must lie in its osculating plane, i.e. the normals to surface S at the points of curve q must lie in plane Q .

Example 3. Consider a spherical surface. Cut this surface by plane Q passing through the centre of the sphere. We shall have the so-called great circle on the spherical surface. *The great circle is the geodesic line on the spherical surface.*

Indeed, the radii of the sphere are the normals to the spherical surface. At the points of the great circle the radii lie in the plane of this circle. Thus we have the case of a plane curve on a surface at the points of which the normals to the surface lie in the plane of this curve. And we have just seen that such a plane curve is the geodesic.

If the sphere is cut by plane Q_1 not passing through the centre of the sphere, a small circle will be obtained on the spherical surface. Since at the points of the small circle the normals to the spherical surface, i.e. the radii of the sphere, do not lie in the plane of this circle, the small circle is not a geodesic on the spherical surface.

A rubber thread tightly stretched along the arc of the great circle will be in equilibrium. If, on the other hand, it is stretched along the arc of the small circle, it will slip off this arc, since it will not be in a state of equilibrium on the arc.

Johann Bernoulli's theorem. *The shortest of all the lines connecting two points on a surface is the arc of the geodesic line.*

We have already obtained the proof of Bernoulli's theorem. Indeed, on the one hand, we have proved that the lines along which threads stretched on a surface in a state of equilibrium lie are geodesics. On the other hand, we know that a rubber thread fixed on a surface at points A and B and situated along the shortest line connecting these points is in a state of equilibrium.*

NOTE. Draw great circle q through two points A and B on a spherical surface. Points A and B divide it into two arcs (Fig. III.3):

\widehat{AMB} and \widehat{ANB} . Both arcs are geodesics connecting points A and B .

If \widehat{AMB} is shorter than \widehat{ANB} , then it is obvious that \widehat{AMB} is the shortest arc on the spherical surface connecting points A and B . Arc \widehat{ANB} , on the other hand, though being a geodesic, will not be the shortest arc on the spherical surface connecting points A and B . A rub-

* A number of other elementary proofs of this theorem are given in the book by M. Ya. Vygodsky *Differentsial'naya Geometriya* (Differential Geometry), Moscow-Leningrad, Gostekhizdat (1949).

ber thread stretched on the spherical surface along any of these arcs will be in a state of equilibrium. But whereas the thread stretched along arc \widehat{AMB} is in a state of *stable* equilibrium, the thread stretched

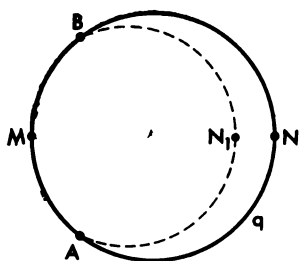


Fig. III.3

along arc \widehat{ANB} is in a state of *unstable* equilibrium. If the thread is taken away from position \widehat{ANB} so that it assumes the form of curve $\widehat{AN_1B}$ (see Fig. III.3) close to \widehat{ANB} but shorter than \widehat{ANB} , it will slip on the spherical surface moving away from position \widehat{ANB} .

Thus we see that *the property of being a geodesic is a necessary but not sufficient condition for the line to be the shortest.*

However, it can be proved that a sufficiently minor arc of a geodesic is always the shortest line.

A geodesic line can be defined as a line whose sufficiently minor arcs are the shortest.

3. The construction of a geodesic line. Move a knife edge on some surface S . Each instant the knife edge will touch the surface at some point A (Fig. III.4). The knife is held so that all the time the normal



- Fig. III.4

to the surface at the point of its contact with the knife edge passes through the knife plane. Line q that will be scratched by the knife edge on surface S will be the geodesic line. Indeed, take minor arc \widehat{BC} of curve q scratched by the knife and point A on it. We can approximately consider that arc \widehat{BC} lies in the knife plane at the instant when the knife edge contacts the surface at point A . So the knife plane at the instant of contact of the knife edge with the surface at point A is the osculating plane of curve q at point A . But we know from the previous discussion that if the osculating plane of curve q continuously passes through the normal to the surface, curve q is the geodesic. Hence, curve q is the geodesic line on the surface.

Let us consider one more problem concerning an arbitrary surface: how to roll out a narrow strip cut out of the surface on a plane and, vice versa, to wrap a narrow strip on the surface. It is now necessary to strictly define what is meant by this.

Given curve q on a surface, encircle it by a narrow strip (Fig. III.5). Generally speaking, this strip cannot be rolled out on a plane so as not to distort the lengths of curves lying on the strip. However, the distortions are the more insignificant the narrower the strip.*

If we lay out a narrow strip onto a plane, the shortest line of the strip connecting two points will pass into an arc whose property will be similar to that on a flat strip, i.e. it will be a straight-line segment. Vice versa, a straight-line segment on a narrow strip which is wrapped

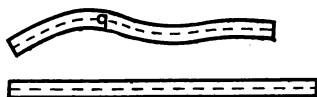


Fig. III.5

on a surface will pass into the shortest arc on the surface, i.e. into a geodesic arc. Therefore, a narrow strip (a band whose width is very small in comparison with its length) encircling a straight-line segment will lie on a surface so that the straight-line segment will pass into the geodesic arc. Our narrow band will lie on the surface along the geodesic line. That is why placing long narrow bands on a surface one can get an idea about the course of the geodesics on the surface.

III.2. Additional Remarks Concerning Geodesic Lines

1. **The symmetry plane.** A few more examples of geodesic lines will be given. First let us recall one definition: *two*

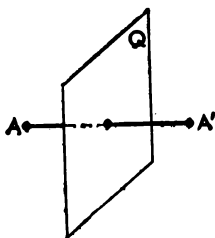


Fig. III.6

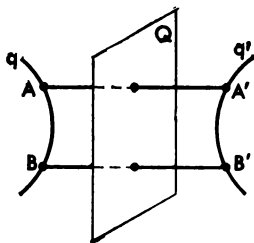


Fig. III.7

points A and A' are symmetrical relative to plane Q if they lie on different sides of plane Q at equal distances from it and on one perpendicular to the plane (Fig. III.6).

* In the language of infinitesimal, changes in the length of the curves will be infinitesimal values of the higher order as compared with the strip width.

Two figures q and q' are symmetric with respect to plane Q if to each point A of figure q there corresponds point A' of figure q' symmetric to A with respect to Q , and vice versa (Fig. III.7).

Plane Q is called the *symmetry plane* of surface S if it divides S into two symmetric parts relative to Q .

Examples. For a spherical surface any plane passing through the centre of the sphere will be the symmetry plane.

For the surfaces of a circular cone and a cylinder the planes passing through the axes will be the symmetry planes.

For a limited circular cylinder a plane perpendicular to the axis of the cylinder and dividing the height into two equal parts will be the symmetry plane.

For an infinitely long cylinder (i.e. a cylinder whose generatrices are infinite straight lines) any plane perpendicular to the axis will be the symmetry plane.

Theorem. Let surface S have symmetry plane Q intersecting S along line q . Line q is the geodesic line of the surface*.

From the premise, line q is in plane Q . Plane line q (see Example 2 of the previous section) will be the geodesic if at any point of curve q the normal to surface S lies in plane Q .

Let point A be an arbitrary point of curve q (Fig. III. 8). Let us prove that the normal to surface S at point A lies in plane Q . Assume the opposite: that normal AB to surface S at point A does not lie in plane Q . Denote by AB' the straight line symmetric to AB with respect to Q . Since AB itself does not lie in Q , AB is different from AB' . But Q is the symmetry plane for the surface, and if AB is the normal to S at point A , straight line AB' symmetric to it is also the normal to S at point A . So surface S has two normals at point A , which is impossible. The contradiction we arrived at proves that the normal to S at any point A of a curve lies in plane Q . The theorem is proved.

2. Closed geodesic lines. If a loop made from a rubber thread is put on surface S so as the thread is in a state of

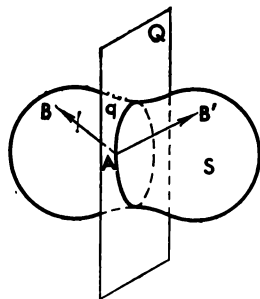


Fig. III.8

* We have agreed to deal only with smooth surfaces.

equilibrium, it will assume the shape of a closed line q . This line q is the *geodesic*. Moreover, it is a *closed* geodesic. Thus a rubber loop on a spherical surface will be in a state of equilibrium if it is stretched along a great circle. The

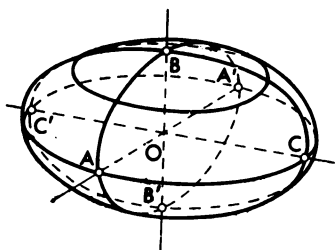


Fig. III.9

An ellipsoid with three axes AA' , BB' and CC' (Fig. III.9) different in length has three symmetry planes each of which passes through two axes of the ellipsoid. These three planes intersect the ellipsoid along three ellipses E_1 , E_2 and E_3 , which are closed geodesics.

It can be proved that *on any closed surface there are at least three closed geodesic lines.**

3. Hertz's principle. A point moving on a plane by inertia follows a straight line (Newton's first law of motion).

A point moving on a surface without external forces acting on it follows a geodesic line.

This is *Hertz's principle*. For example, a point on a spherical surface moves along a great circle if it is not acted upon by external forces. A point on a cylindrical surface moves along a helix under the same circumstances.

Indeed, the acceleration with which a point moves along curve q can be resolved into a *tangential* acceleration (directed along the tangent to q) and a *normal* acceleration (directed along the principal normal to q). But if the point moves unaffected by external forces in the direction of curve q located on surface S , then only the surface reaction acts upon the points, the force of this reaction being directed

* The proof of this non-elementary theorem is given in the article by L. A. Lyusternik and L. Shnirelman "Topological Methods in Variational Problems and Their Application to Differential Geometry of Surfaces", *Uspekhi Matematicheskikh Nauk*, 2, No. 1, 17 (1947).

along the normal to the surface. Since the direction of force coincides with the direction of acceleration, at each moment the direction of acceleration of the point must coincide with the direction of the normal to the surface. The normal to a surface at some point of a curve is perpendicular to the tangent to curve q at the same point. Since the acceleration is directed along the normal to the surface, i.e. perpendicular to the tangent to q , the tangential acceleration is zero. Hence, our point is only under the effect of the normal acceleration directed along the principal normal to q . The direction of acceleration is at the same time the direction of the principal normal to curve q and the normal to surface S . It means that these directions coincide at any point of curve q and, whence, that curve q is a geodesic line on surface S .

4. Geodesic lines on a surface with an edge. Consider surface S consisting of two smooth surfaces S_1 and S_2 adjoining each other along curve s which we shall agree to

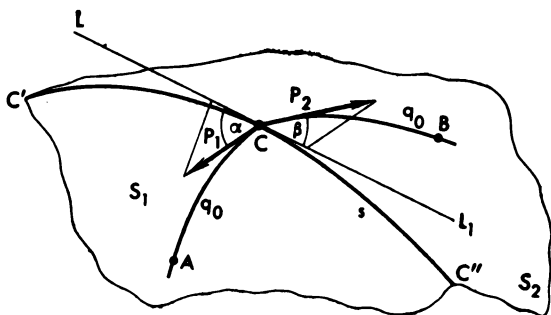


Fig. III.10

call the *edge* of surface S (an example of such a surface may be a dihedral angle). On surface S there are two points A and B lying, respectively, on S_1 and S_2 (Fig. III.10). Let $q_0 = ACB$ be the position of equilibrium of an elastic thread on surface S . Point C belongs to edge s , and arcs \widehat{AC} and \widehat{CB} of curve q_0 , respectively, to parts S_1 and S_2 . It is obvious that \widehat{AC} is the geodesic on S_1 , and \widehat{CB} the geodesic on S_2 . Let us find the condition of equilibrium at the point of inflection C by using the method employed in Sec. III.1. Curve q_0 is the position of equilibrium of the flexible thread fixed at points A and B on surface S .

Denote by α the angle between arc \widehat{AC} and part CC' of edge s , and by β the angle between part CC'' of edge s and arc \widehat{CB} (i.e. between their tangents). Point C is under the effect of the tensile strengths, viz. P_1 directed along the tangent to arc \widehat{CA} , and P_2 directed along the tangent to arc \widehat{CB} . Each of these forces is equal to T . The projections of these forces on tangent LL_1 to edge s at point C are equal, respectively, to $T \cos \alpha$ and $T \cos \beta$ and are oppositely directed. The condition of equilibrium

$$T \cos \alpha = T \cos \beta$$

yields

$$\alpha = \beta \quad (\text{III.1})$$

The angles formed by arcs \widehat{AC} and \widehat{CB} with edge s at the inflection point are equal to each other.

Line q_0 is naturally called the *geodesic* on surface S .

If surface S consists of several smooth parts divided by edges

$$s_1, s_2, \dots, s_n$$

the geodesic lines (lines of equilibrium of an elastic thread) on such surfaces will consist of geodesic arcs joining on the edges

$$s_1, s_2, \dots, s_n$$

so condition (III.1) holds at each point of the joint.

The shortest lines on surface S are the geodesics. The property of the shortest lines on polyhedral surfaces deduced in Sec. I.1 is a particular case of the property of the geodesic (and shortest) lines on a surface with edges.

This property of geodesics on such surfaces can also be derived from Hertz's principle.

III.3. Geodesic Lines on Surfaces of Revolution

1. A surface of revolution. Rotate plane curve q about straight line AB lying in one plane with q (Fig. III.11). In revolving q about AB surface S is formed, which is called the *surface of revolution*. Any plane Q passing through axis of revolution AB intersects S along two curves q and q' . These curves are called *meridians*. They are obtained from curve q by turning it through a certain angle about the axis of revolution. Each plane perpendicular to the axis intersects S along a circle called the *parallel*.

Theorem 1. *All the meridians of a surface of revolution are geodesic curves.*

Consider meridians q and q' formed by the intersection of the surface of revolution with plane Q passing through axis AB . Plane Q is the symmetry plane of surface of revolution S . Consequently, it crosses surface S along geodesic curves. Thus lines q and q' are geodesics.

Example. Revolve ellipse E about its axis (Fig. III.12). We shall obtain the so-called spheroid. Its meridians are but ellipses equal to E . These ellipses are geodesic ones.

NOTE. On the surface of a cylinder all parallels are geodesics; on the surface of a sphere of all the parallels only the equator is the geodesic; on the surface of a cone no parallel is a geodesic.

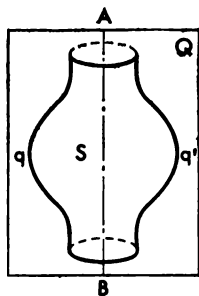


Fig. III.11

2. Clairaut's theorem. Consider geodesic line q on surface of revolution S . Let A represent an arbitrary point of geodesic q , r its distance from the axis of revolution (the paral-

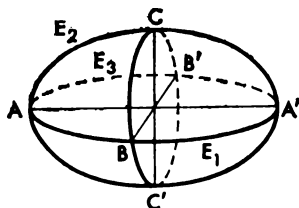


Fig. III.12

lel radius), α the angle between geodesic q and the meridian at point A .

Theorem 2 (Clairaut's theorem). *The expression $r \sin \alpha$ has a constant value in all points of geodesic q :*

$$r \sin \alpha = c = \text{const} \quad (\text{III.2})$$

If the angle between the geodesic and the parallel is denoted by β , formula (III.2) will assume the form

$$r \cos \beta = \text{const}$$

The particular case of Clairaut's theorem for conical and cylindrical surfaces has already been proved (see Sec. I.3, Item 4).

Consider surface S_n formed by the rotation of broken line $A_0A_1 \dots A_n$ about axis L . Surface S_n consists of n surfaces s_1, s_2, \dots, s_n formed by the rotation of the respective sides $A_0A_1, A_1A_2, \dots, A_{n-1}A_n$. These surfaces are separated from one another by edges t_1, t_2, \dots, t_{n-1} , i.e. by parallels obtained through the rotation of the broken line vertices A_1, A_2, \dots, A_{n-1} .

Consider two points A and B on surface S_n and geodesic q_0 connecting them. Owing to the result in Item 4 of Sec. III.2, geodesic q_0 consists of geodesic arcs on the surfaces of truncated cones or cylinders

$$s_1, s_2, \dots, s_n$$

joining by the edges

$$t_1, t_2, \dots, t_{n-1}$$

where the angles formed by each of the joining geodesic arcs with the edge are all the same. When angle β of curve q_0 with the parallel moves along q_0 , it changes gradually, without jumps (a jump in the change of the angle could occur if the parallel becomes one of the edges, which is impossible according to previous result). Therefore the quantity $r \cos \beta$ changes gradually, or without jumps.

Let us see what will happen to the quantity $r \cos \beta$ while moving along q_0 . While moving along one of the surfaces

$$s_0, s_1, \dots, s_n$$

the expression $r \cos \beta$ will remain constant (by virtue of the particular case of Clairaut's theorem already proved). When passing across one of the edges

$$t_1, t_2, \dots, t_{n-1}$$

this expression will change gradually. So it remains constant along the whole of q_0 . Thus for all points of geodesic q_0 the relation

$$r \cos \beta = \text{const}$$

will hold.

An arbitrary plane curve m can be considered as the limiting curve for inscribed polygons m_n when the number of the sides grows unlimitedly and the length of the maximum side tends to zero. Surface S formed by the rotation of m about some axis is the limiting one for surfaces S_n formed by the rotation of m_n about the same axis. For the shortest lines on surfaces S_n Clairaut's theorem is valid. From this we can draw a conclusion that the same is true for the shortest lines on surface S .

LECTURE 2

CHAPTER IV Problems Associated with the Potential Energy of Stretched Threads

IV.1. Motion of Lines that Does Not Change Their Length

1. The potential energy of a flexible thread. Assume that a flexible thread possesses equal tension T in all of its points



Fig. IV.1

and that this tension remains in spite of changes in the thread's length. We want to determine the potential energy of the thread.

Let $q = \widehat{ABC}$ be a smooth curve of length l consisting of arcs \widehat{AB} of length l_0 and \widehat{BC} of length $(l - l_0)$ (Fig. IV.1). Let the thread which occupied position \widehat{AB} pass, twisting along curve q , into position $\widehat{A'E'E''C}$ with point A being fixed and point B following line \widehat{BC} of length $(l - l_0)$. Consider the work carried out by the tensile strengths.

The tensile strengths at point B have carried out work equal to $T(l - l_0)$.

The work of tensile strengths acting on minor arc $\widehat{E'E''}$ of curve q is zero. Indeed, the resultant of these forces is directed along the normal to curve q , whereas arc $\widehat{E'E''}$ slides along curve q .

Thus the total work of the tensile strengths with the thread moving as mentioned above amounts to the work of the force applied to end B , i.e. equals

$$T(l - l_0) = Tl - Tl_0$$

Let the potential energy of the thread in position \widehat{AB} be V_0 , and the potential energy when the thread occupied position \widehat{ABC} be V . The potential energy increment $V - V_0$ is equal to the work carried out, i.e.

$$V - V_0 = Tl - Tl_0$$

or

$$V - Tl = V_0 - Tl_0 \quad (\text{IV.1})$$

Assume that the potential energy also tends to zero when the length of the thread tends to zero, then when $l_0 \rightarrow 0$ we have $V_0 \rightarrow 0$; consequently, $(V_0 - Tl_0) \rightarrow 0$. Passing over to the limit in the right-hand side of equation (IV.1) as $l_0 \rightarrow 0$, we shall have

$$V - Tl = 0$$

whence

$$V = Tl \quad (\text{IV.2})$$

The potential energy of a flexible thread is equal to its length multiplied by tensile strength.

COROLLARY. If the work of the tensile strengths is zero, the length of the thread has not changed. Indeed, under such conditions the potential energy of the thread, the energy to which the thread length is proportional, has not changed.

Note that if straight-line segment AB remains straight while being in motion, the total work of the tensile strengths amounts to the work of the tensile strengths at the ends of this segment.

The work of the thread retaining the shape of broken line ACB amounts to the work of the tensile strengths at ends A and B of the broken line and at its vertex C .

2. Parallel lines. Two lines with common normals are called *parallel*. The simplest examples of parallel lines are given by parallel straight lines and concentric circles.

Theorem 1. *Segments of common normals between parallel lines q and q_1 are of equal lengths.*

Let AB , a common normal to curves q and q_1 , move from position A_0B_0 to position A_1B_1 remaining common to them all the time (Fig. IV.2).

The work of tensile strengths in such motion is zero. Indeed, at end A the tensile strength is directed along the normal to the curve. Therefore, when this end moves along curve q , the work of the tensile strengths is zero. Similarly,

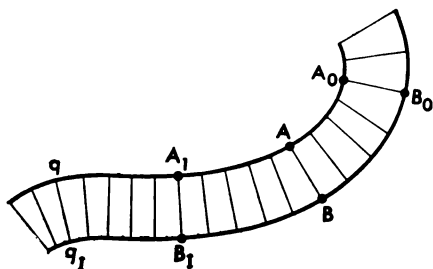


Fig. IV.2

at end B moving along curve q_1 the work of the tensile strengths is zero. So, during such motion of the common normal the work of the tensile strengths is zero. By virtue of the stated corollary, while being in motion length l of the common normal does not change:

$$l(A_0B_0) = l(A_1B_1)$$

3. Normals to an ellipse and a parabola. An *ellipse* is a locus of points B whose sum of distances from the given points F and F_1 is a constant:

$$FB + F_1B = 2a \quad (\text{IV.3})$$

where a is a constant.

Points F and F_1 are called the *foci*, and segments FB and F_1B *radius vectors* of an ellipse.

Theorem 2. The normal to an ellipse at any point B is bisector BD of angle FBF_1 formed by the radius vectors (Fig. IV.3).

Let an elastic thread having the shape of broken line FBF_1 be fixed at points F and F_1 . If the thread is displaced by moving point B along the ellipse, then by virtue of equation (IV.3) the length of the thread does not change. Thus the work of the tensile strengths is zero all the time. The

work amounts to the work of forces at point B . At this point two equal tensile strengths are applied acting along BF and BF_1 . Their resultant P is directed along bisector BD

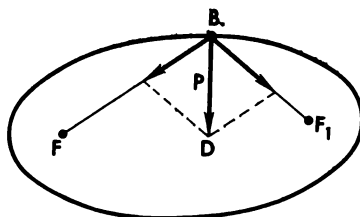


Fig. IV.3

of angle FBF_1 . Since in moving point B along the ellipse work P is zero all the time, P is always directed along the normal to the ellipse. Hence, the normal to the ellipse at any point B coincides with the bisector of angle FBF_1 .

A *parabola* is a locus of points B whose distances from the given point F and from the given straight line d are equal:

$$FB = BC \quad (\text{IV.4})$$

where BC is a perpendicular dropped from B to straight line d (Fig. IV.4).

Point F is called the *focus*, straight line d the *directrix*, and straight line LL perpen-

dicular to d and passing through the focus the *axis* of a parabola.

Draw straight line d_1 parallel to d so that focus F and directrix d appear at one side from d_1 . Denote by a the distance between parallel straight lines d and d_1 . Draw a common perpendicular CC_1 to straight lines d and d_1 through point B of the parabola (CC_1 is parallel to axis LL). We have

$$CC_1 = CB + BC_1 = a$$

where a is a constant equal to the distance between parallel straight lines d and d_1 .

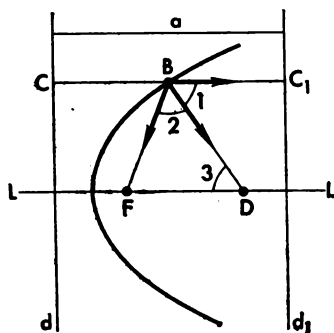


Fig. IV.4

By virtue of equation (IV.4),

$$FB + BC_1 = a \quad (\text{IV.5})$$

Now it is easy to prove the following

Theorem 3. *The normal at an arbitrary point B of the parabola is the bisector of angle FBC formed by radius vector FB and straight line BC_1 parallel to axis LL .*

Consider a thread in the form of broken line FBC_1 whose end F is fixed, end C_1 slides along straight line d_1 so that BC_1 remains perpendicular to d_1 , and point B slides along the parabola.

The length of the thread, as is evident from equation (IV.5), remains unchanged. Hence, the total work of the tensile strengths is zero. This work is the sum of the work of the tensile strengths at points C_1 and B . The work of the tensile strengths at point C_1 is zero because the direction of this force (along segment C_1B) is normal to d_1 along which point C_1 moves. Hence, the work of the tensile strengths at point B is also zero. Repeating the reasoning employed in examining the ellipse we prove the theorem.*

NOTE. From Theorem 3 there follows the rule for constructing a normal to a parabola. Draw segment FD equal to radius vector FB of the parabola along axis LL . Straight line BD is the normal to the parabola.

Indeed, in Fig. IV.4 angles $\angle 1$ and $\angle 3$ are equal as alternate-interior angles at parallels LL and CC_1 and secant BD ; angles $\angle 3$ and $\angle 2$ are equal because FBD is an isosceles triangle. Thus $\angle 2 = \angle 1$, i.e. BD is the bisector of angle FBC , and by virtue of Theorem 3 it is the normal to the parabola at point B .

4. Geodesic tangents and normals. If geodesic arc \widehat{AB} moves over a surface, the work is performed only by the tensile strengths acting at the ends A and B of the arc. Indeed, the resultant of forces acting on any minor inner part of arc \widehat{AB} is directed along the normal to the surface, and so while moving along the surface its work is zero.

* True, the theorem is thus proved for the points located in the part of the parabola to the left from straight line d_1 . But since the position of the straight line (parallel to d) is arbitrary, the theorem holds for all points of the parabola.

Geodesic line r having a common tangent with curve q at point B is called the *geodesic tangent* to curve q on a surface at point B ; geodesic curve s orthogonal to q at point B is called the *geodesic normal* to curve q at point B (Fig. IV.5).

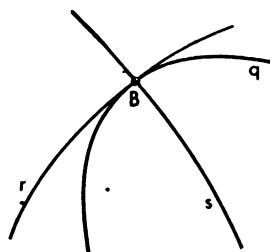


Fig. IV.5

Theorem 1 on common normals is generalized for the case of geodesic normals.

Theorem 4. Let two curves q and q_1 on a surface have all geodesic normals common. Segments of the common geodesic normals between q and q_1 have the same length (Fig. IV.6).

Example. Segments of meridians on the surface of a sphere between two parallels are of the same length.

In proving Theorem 4 we repeat the proof of Theorem 1.

5. Geodesic circles. Draw all kinds of geodesic arcs \widehat{AB} of the same length from point A of the surface. Locus q of

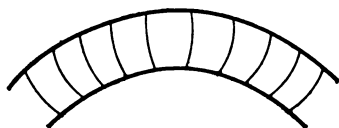


Fig. IV.6

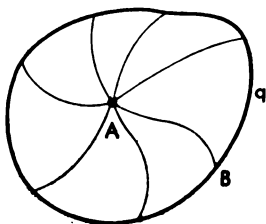


Fig. IV.7

their ends B is called the *geodesic circle*; geodesic arcs \widehat{AB} are called the *geodesic radii* (Fig. IV.7).

Each geodesic radius \widehat{AB} is the geodesic normal to the geodesic circle at point B .

Indeed, suppose elastic thread \widehat{AB} is fixed at end A and having the shape of a geodesic radius moves so that its end B circumscribes geodesic circle q . Since the length of geodesic arc \widehat{AB} does not change, the work of the tensile strengths is zero. This work amounts to the work of the tensile strengths at end B . So the work of the tensile strengths at B equals

zero all the time. The tensile strengths are directed along the normal to the line of displacement q . And since at point B they are directed tangentially to radius \widehat{AB} , we have proved the theorem.

IV.2. Evolutes and Involutives

Consider plane curve q , a family of normals drawn from various points of this curve, and *envelope* s of these normals (i.e. curve s touching these normals). Envelope s is called the *evolute* of curve q , and curve q , intersecting orthogonally all the tangents to evolute s , is called its *involute* (Fig. IV.8).

Each point B of the evolute is the point of intersection of normal AB to the involute and an infinitely close normal $A'B'$, i.e. point B is the centre of curvature for curve q at point A (see Sec. II.3). Evolute s of curve q can be termed as the *locus of the centres of curvature of this curve*.

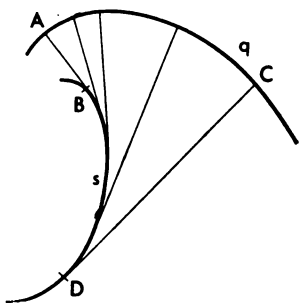


Fig. IV.8

Suppose that an elastic thread has the form of curve r , consisting of a segment of normal AB to the involute and of arc \widehat{BD} of evolute s (see Fig. IV.8). Moving along this curve from A to D we have at point B a smooth transition from segment AB to arc \widehat{BD} . Therefore, the elastic thread in position $r = \widehat{ABD}$ is in equilibrium. Move thread r so that its end A follows an involute, and point B an evolute. Then AB preserves being the normal to the involute, and the remaining part of the thread \widehat{BD} adjoins curve s . The work of the tensile strengths acting at the points of normal AB equals the work at points A and B . But this work is zero at point A because the tensile strengths act along the normal to curve q , along which end A slips. The tensile strengths applied to point B counterbalance, and their work at each given moment is zero. Finally, the work is also zero on part \widehat{BD} of thread r which is not moving at the given moment. Thus the work is zero at any moment. During the motion the po-

tential energy of thread r remains unchanged, hence, so does the length of thread r .

If \widehat{ABD} is the initial position of thread r , and segment CD its final position, the length of \widehat{ABD} is equal to CD :

$$l(\widehat{ABD}) = l(CD)$$

But

$$l(\widehat{ABD}) = l(AB) + l(\widehat{BD})$$

or

$$l(CD) = l(AB) + l(\widehat{BD})$$

whence

$$l(\widehat{BD}) = l(CD) - l(AB)$$

This constitutes the proof of the following

Theorem. *If normals AB and CD are drawn from two points A and C of an involute to their points of contact B and D with an evolute, the difference in the lengths of these segments of the normals is equal to the length of the arc of evolute \widehat{BD} enclosed between them.*

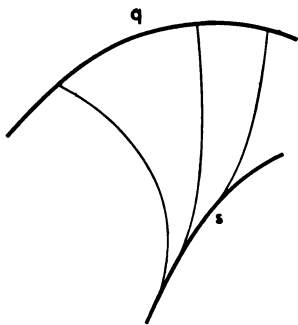


Fig. IV.9

If a family of geodesic normals is drawn to curve q on a surface (Fig. IV.9), their envelope s is called the *geodesic evolute* of curve q , and curve q the *geodesic involute* of curve s . The theorem holds true if the words “normals”, “evolutes” and “invo-

lutes” are understood as geodesic normals, geodesic evolutes and geodesic involutes. The reader may satisfy himself that the proof will be valid in this case too.

IV.3. Problems of Equilibrium of a System of Elastic Threads

1. The Dirichlet principle. *For a mechanical system the position of a minimum of its potential energy is an equilibrium position.* Indeed, if an immovable mechanical system is shifted from position S of a minimum of the potential energy, its potential energy can only grow. Hence, in accordance with the law of conservation of energy, its kinetic energy

can only diminish. Therefore, if in position S the system is immovable, i.e. the kinetic energy is zero, then when shifted it cannot acquire any positive kinetic energy.

Example. For an elastic thread the potential energy is proportional to its length. Therefore, the position in which the thread is of the minimum length is the position of equilibrium. This circumstance has already been used.

Consider the following problems of finding the equilibrium position of a system of *several* threads (the second of the problems will be important for subsequent discussion).

2. A problem of the minimum sum of lengths. Given points B_1, B_2, \dots, B_n on a plane. Find point A with the minimum sum of distances from the given points. Consider n elastic threads AB_1, AB_2, \dots, AB_n which have one common end A (for instance, the threads are tied up at point A), and the other ends are fixed, respectively, at points B_1, B_2, \dots, B_n . The potential energy of this system of threads is proportional to the sum of lengths of threads AB_1, AB_2, \dots, AB_n . The minimum sum of lengths of the threads, i.e. the minimum potential energy, corresponds to the equilibrium position of the system. In this position each thread becomes a straight-line segment and the sum of lengths of the threads is a minimum. Let A_0 be the position of point A in this state of equilibrium (Fig. IV.10). Point A_0 is under the action of n tensile strengths directed along $A_0B_1, A_0B_2, \dots, A_0B_n$. These n forces are equalized. So at point A_0 , for which the sum of distances to points B_1, B_2, \dots, B_n is the minimum, the resultant of n equal forces acting along $A_0B_1, A_0B_2, \dots, A_0B_n$ is zero.*

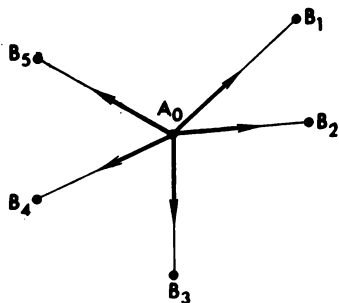


Fig. IV.10

* M. Ya. Vygodsky has shown that this statement must be specified. It is correct if point A , for which the sum of lengths AB_1, AB_2, \dots, AB_n is minimal, does not coincide with either one of points B_1, B_2, \dots, B_n .

For instance, in the case of three points B_1, B_2 and B_3 , point A lies inside triangle $B_1B_2B_3$ if neither of its angles is more than 120° . If, on the other hand, one of them, for instance at vertex B_1 , is equal to or is more than 120° , point A coincides with this vertex.

This point A_0 can be found mechanically in the following manner. On a horizontal plate n holes are drilled at points B_1, B_2, \dots, B_n (Fig. IV.11). Secure n strings at one point over the plate, pass these strings through the holes and suspend weights of equal mass from them. The system of strings with the weights will come to equilibrium, and the common point of the strings in the state of equilibrium will be the sought point A_0 . Indeed, this point is under the

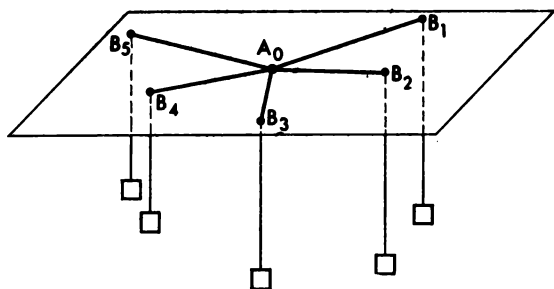


Fig. IV.11

action of n equal tensile strengths of the strings, the forces being directed to holes B_0, B_1, \dots, B_n (each of the forces is equal to the mass of the weight suspended from the string). These n equal tensile strengths are equalized.

The following specific problem amounts to that discussed above. Given n points B_1, B_2, \dots, B_n . A storehouse is to be built at point A with straight roads AB_1, AB_2, \dots, AB_n from it. Find the most favourable position of the storehouse so that the sum of the lengths of roads AB_1, AB_2, \dots, AB_n is minimal.

Sometimes this problem is complicated. Let freight trafics from storehouse A to points B_1, B_2, \dots, B_n be respectively proportional to q_1, q_2, \dots, q_n . Find the position of point A for which the sum

$$S = q_1 \overline{AB_1} + q_2 \overline{AB_2} + \dots + q_n \overline{AB_n}$$

is the minimum (i.e. for which the number of ton-kilometres in carrying cargoes by routes AB_1, AB_2, \dots, AB_n is the minimum).

The problem is solved similarly to the previous one (which is its particular case with $q_1 = q_2 = \dots = q_n$).

We want to find the position of equilibrium of a system of n threads AB_1, AB_2, \dots, AB_n fixed at points B_1, B_2, \dots, B_n and having a common point A . But threads AB_1, AB_2, \dots, AB_n are stretched differently in proportion to q_1, q_2, \dots, q_n , or equal to $q_1 T, q_2 T, \dots, q_n T$, respectively. The potential energy of threads AB_1, AB_2, \dots, AB_n is equal, respectively, to $q_1 T \overline{AB_1}, q_2 T \overline{AB_2}, \dots, q_n T \overline{AB_n}$. The total potential energy of the system is equal to

$$V = T (q_1 \overline{AB_1} + q_2 \overline{AB_2} + \dots + q_n \overline{AB_n}) = TS \quad (\text{IV.6})$$

The position with the minimum value of V , i.e. the minimum value of sum S , is the position of equilibrium of the system. Each line $AB_i, i = 1, 2, \dots, n$ will then become a straight-line segment. The common point $A = A_0$ of these threads is in equilibrium under the effect of n tensile strengths directed along segments $A_0 B_1, A_0 B_2, \dots, A_0 B_n$ and is proportional to q_1, q_2, \dots, q_n .

The mechanical method of finding the sought point A_0 stated above remains valid. However, in this case the weights attached to the ends of the strings threaded through the holes at points B_1, B_2, \dots, B_n should be proportional to q_1, q_2, \dots, q_n .

3. A problem on the equilibrium of a system of two threads. Consider a flexible non-homogeneous thread of shape $q = \widehat{ACB}$ (Fig. IV.12) whose ends A and B are fixed and whose point C moves along curve s . The tension of part \widehat{AC} is equal to T_1 and that of part \widehat{CB} to T_2 . The potential energy $V(q)$ of the thread is

$$V(q) = V(\widehat{AC}) + V(\widehat{CB})$$

Since

$$V(\widehat{AC}) = T_1 l(\widehat{AC})$$

and

$$V(\widehat{CB}) = T_2 l(\widehat{CB})$$

we have

$$V(q) = T_1 l(\widehat{AC}) + T_2 l(\widehat{CB}) \quad (\text{IV.7})$$

Let thread q have the minimum potential energy in position q_0 . By virtue of the Dirichlet principle, the thread in

position q_0 is in equilibrium. Let C_0 be the point of intersection of q_0 and s .

No doubt that each of the parts $\widehat{AC_0}$ and $\widehat{C_0B}$ of line q_0 is a straight-line segment. Consider the conditions of equilibrium at point C_0 .

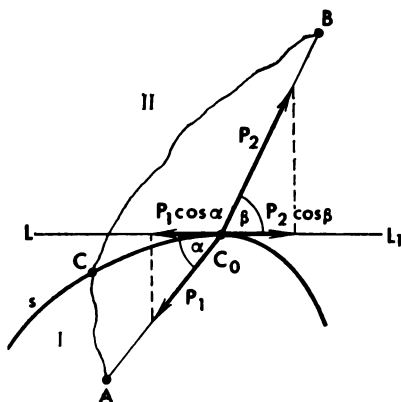


Fig. IV.12

Applied to this point are the following tensile strengths: force P_1 directed along C_0A and equal to T_1 , and force P_2 directed along C_0B and equal to T_2 . Draw tangent LL_1 to curve s at point C_0 . Denote the angles

$$\left. \begin{aligned} \angle AC_0L &= \alpha \\ \angle L_1C_0B &= \beta \end{aligned} \right\} \quad (\text{IV.8})$$

The tangential component of force P_1 equals $P_1 \cos \alpha = T_1 \cos \alpha$ and is directed along C_0L ; the tangential component of force P_2 equals $P_2 \cos \beta = T_2 \cos \beta$ and is directed along C_0L_1 . Point C_0 is in equilibrium if both tangential components are equalized, i.e. if

$$T_1 \cos \alpha = T_2 \cos \beta \quad (\text{IV.9})$$

So line q_0 is broken line AC_0B with the vertex at C_0 on boundary curve s , where condition (IV.9) is fulfilled.

CHAPTER V

The Isoperimetric Problem

V.I. Curvature and Geodesic Curvature

1. Curvatures. The value $1/R$, the reverse of radius R of a circle, is called the *curvature* of the circle. This notion can be mechanically illustrated with the aid of a stretched thread.

Given arc \widehat{AB} of a circle of radius R with the centre at O . Assume that the arc is formed by an elastic thread with equal tensile strengths T_1 and T_2 applied to its ends and directed along the tangents as illustrated in Fig. V.1.

The resultant T_0 of forces T_1 and T_2 is directed along the bisector of the angle between the directions of forces T_1 and T_2 , i.e. along the radius dividing arc \widehat{AB} in two. If this arc is α radians, its length is equal to $R\alpha$ and the length of the contracting chord is $2R \sin(\alpha/2)$. Since a minor arc can be considered approximately equal to the chord, it follows that $2R \sin(\alpha/2) \approx R\alpha$. Thus with very small magnitudes of angle α we shall have $\sin(\alpha/2) \approx \alpha/2$, i.e. a small angle expressed in radians is approximately equal to its sine.

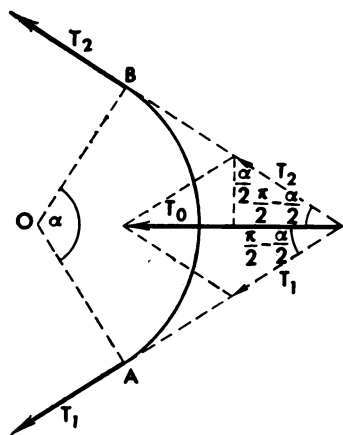


Fig. V.1.

NOTE. More accurately, the ratio of an angle to its sine tends to unity when the angle tends to zero. The proof of this theorem can be found in any course in calculus, as well as in textbooks on trigonometry.

To make our subsequent reasonings more accurate, the concept of *equivalent infinitesimals* should be introduced. An infinitesimal is a variable which tends to zero.

Let quantity β tend to zero together with quantity α (e.g. the length of chord tends to zero simultaneously with the arc it contracts). If the ratio β/α of infinitesimals β and α is also an infinitesimal, β is called the *infinitesimal of higher order* in comparison with α . For example, α^2 is an infinitesimal of higher order as compared with α .

Two infinitesimals α and γ are called *equivalent* if their ratio tends to unity:

$$\lim_{\alpha \rightarrow 0} \frac{\gamma}{\alpha} = 1 \quad (\text{V.1})$$

For example, a chord contracting an arc is equivalent to the latter.

The difference of two equivalent infinitesimals γ and α is an infinitesimal of higher order in comparison with any one of them. Indeed, it follows from (V.1) that

$$\lim_{\alpha \rightarrow 0} \frac{\gamma - \alpha}{\alpha} = 0 \quad (\text{V.2})$$

For this reason the error due to the replacement of an infinitesimal by its equivalent is an infinitesimal of higher order. For example, the difference in the length of an infinitesimal arc and a chord contracting it is an infinitesimal of higher order. The error ensuing from equating the arc and the chord is an infinitesimal of higher order as compared with the related quantities.

The equivalency of values α and γ is expressed by using the notation $\alpha \approx \gamma$.

An example of equivalent quantities is $\sin \alpha \approx \alpha$ with an infinitesimal α [this is a symbolic expression of the equality $\lim_{\alpha \rightarrow 0} (\sin \alpha / \alpha) = 1$].

Denote angle AOB (see Fig. V.1) measured in radians by α . Then the angle between the directions of forces T_1 and T_2 is equal to $\pi - \alpha$, and the angle between their directions and the direction of the resultant T_0 is $(\pi/2) - (\alpha/2)$.

As is seen from the drawing, $T_0 = 2T \sin (\alpha/2)$, where T is the common quantity of forces T_1 and T_2 .

If we denote the length of arc \widehat{AB} by s , its quantity α in radians will be given as $\alpha = s/R$.

Hence,

$$T_0 = 2T \sin \frac{s}{2R}$$

If arc s is minor, then

$$\sin \frac{s}{2R} \approx \frac{s}{2R}$$

and

$$T_0 = T \frac{s}{R}$$

Consider now an arbitrary curve q . A minor arc of length s of this curve containing point A may be considered as the arc of a circle (whose radius R is the radius of curvature at point A). Let line q be an elastic thread whose points are under tension equal to T . Then the arc is under the action of two tensile strengths at the ends whose resultant, by virtue of the said above, is directed along the radius of curvature and is equal to (more accurately, is equivalent to) $T (s/R)$.

The value $1/R$ is called the curvature of the line at point A .

Thus, minor arc \widehat{AB} is under the action of the force which is directed along the principal normal and is proportional to the length of arc s and curvature $1/R$.

2. The geodesic curvature. Consider a minor arc of length s of curve q lying on a surface, and let A be the midpoint of this arc (Fig. V.2). Denote the curvature of the line at

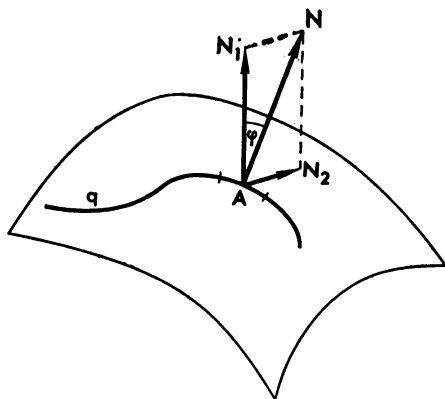


Fig. V.2

point A by $1/R$, and the angle between principal normal AN of curve q at point A and normal AN_1 of the surface by φ . At point A the arc is under the action of a force directed along the principal normal to curve q at point A . The force equals $T (s/R)$. Resolve the force into two forces: one acting along

the normal to the surface (this force is annihilated by the reaction of the surface) and another contacting the surface. This second force will make the arc to slide over the surface. It is equal to (more accurately, is equivalent to)

$$\frac{T s \sin \varphi}{R} = T s \Gamma$$

The value $\Gamma = \sin \varphi / R$ is called the *geodesic curvature* of line q at point A . It determines the intensity of the force acting on the arc of the stretched thread at point A and making the arc to slide over the surface. The force acting on the minor arc of the curve is proportional to the length of arc s and to the geodesic curvature Γ .

For the geodesic line for which $\varphi = 0$ the geodesic curvature is zero. There is no force which makes the arc of the line to slide over the surface, the thread stretched along the geodesic line being in a state of equilibrium.

V.2. An Isoperimetric Problem

1. Measuring the length of an arc of a circle. Given circle q of radius R and arc \widehat{AB} of this circle. Let AB be an arc

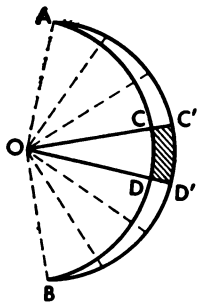


Fig. V.3

that is very close to the latter.* Denote the length of arc \widehat{AB} by l and the length of arc AB by $l + \Delta l$. If arc \widehat{AB} is transformed so that it passes into arc AB , its length l will increase by Δl , and its potential energy will therefore increase by $T\Delta l$. Let us transform \widehat{AB} into AB so that each of its points C shifts along a radius (Fig. V.3). Let minor arc \widehat{CD} (part of \widehat{AB}) pass into another minor arc $C'D'$ (part of AB). Each point of

this arc will shift by segment $\widehat{CC'}$ (in view of the smallness of \widehat{CD} we shall assume the displacement of its

* By saying this we assume that the points of this arc are very close to those of the former arc, and that the curvature of the new arc is close to that of the previous one.

points to be roughly the same). The small area $CC'D'D$ bounded by the arcs and by segments CC' and DD' may be approximately considered as a rectangle. If h is the length of the minor arc \widehat{CD} , area $CC'D'D$ is approximately equal* to hCC' :

$$\text{area } CC'D'D \approx h \cdot CC' \quad (\text{V.3})$$

Note that arc \widehat{CD} is under the action of a force directed along the radius and equal to Th/R , where R is the radius of our circle. The work performed in moving arc \widehat{CD} until it coincides with $C'D'$ is equal to force Th/R multiplied by the path CC' , i.e. $(Th/R) CC'$ or [see formula (V.3)]

$$\frac{Th}{R} CC' = \frac{T}{R} (\text{area } CC'D'D) \quad (\text{V.4})$$

So the work that must be performed in order to displace minor arc \widehat{CD} to a new close position $C'D'$ is equal to (more accurately, is equivalent to) T/R multiplied by area $CC'D'D$ that will be swept by the arc in its motion.

Denote the area enclosed between arcs \widehat{AB} and AB by ΔF . Break up this area into small areas (similar to area $CC'D'D$) by radii drawn from centre O . Arc \widehat{AB} will be also divided into minor arcs. While shifting, each such arc \widehat{CD} will circumscribe a respective area $CC'D'D$ (contained between this arc, arc $C'D'$ and segments of radii CC' and DD'). The work performed in the motion is equal to T/R multiplied by the area circumscribed by this arc. The total work performed in the motion of the entire arc \widehat{AB} to position AB is equal to the sum of elemental works, i.e. to T/R multiplied by the sum of the small areas, i.e. $(T/R) \Delta F$, where ΔF is the area swept by arc \widehat{AB} in its motion.

But the work performed is equal to the increment of the potential energy ΔV while arc \widehat{AB} passes into AB :

$$\Delta V \approx \frac{T}{R} \Delta F \quad (\text{V.5})$$

* In the sense of equivalency.

On the other hand, it follows from formula (IV.2) of Sec. IV.1 that

$$\Delta V = T \Delta l \quad (\text{V.6})$$

where Δl is the increment of the length.

Comparing expressions (V.5) and (V.6), we get

$$\frac{T}{R} \Delta F \approx T \Delta l$$

or

$$\Delta l = \frac{1}{R} \Delta F \quad (\text{V.7})$$

Increment Δl of the length of arc \widehat{AB} is equal to (more accurately, is equivalent to) curvature $1/R$ multiplied by the area contained between arcs \widehat{AB} and AB .*

2. Changing the length of an arc of an arbitrary curve.
If an arbitrary curve is taken instead of a circle, its minor arc \widehat{AB} can be considered as the arc of a circle of radius R , R being the radius of curvature. Formula (V.7) will hold true if $1/R$ is understood to be the curvature of the curve at some midpoint of arc \widehat{AB} .

The same is observed for curves displaced on a surface, the only difference being that curvatures will turn everywhere into geodesic curvatures. Formula (V.7) will take the form

$$\Delta l = \Gamma \Delta F \quad (\text{V.8})$$

where Γ = geodesic curvature

Δl = increment of the length of arc of the curve when replacing it with an arc close to it on the same surface

ΔF = area enclosed between the primary and changed arcs.

In Fig. V.3 area ΔF lies *outside* the circle whose arc is \widehat{AB} . In Fig. V.4 it is *inside* the circle. Let us agree to consider the latter area ΔF as a negative one. Increment Δl of the length of arc of the circle will also be negative (since we have the arc's shortening and not its lengthening).

* All the equations are accurate to infinitesimals of higher order in comparison with Δl .

3. An isoperimetric problem. Let us consider the following problem. *Of all closed curves bounding the area of a given magnitude F find the one which is the shortest.*

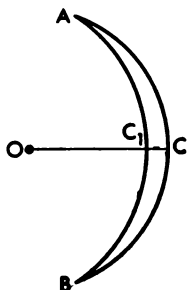


Fig. V.4

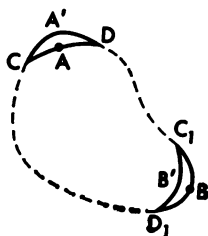


Fig. V.5

Assume that such a curve does exist. Let us prove that it is a circle.

Note that a curve of a constant curvature (i.e. a curve which has the same curvature $1/R$ at all of its points) is a circle.

A proof of this fact (a proof which does not claim to be a very strict one) is as follows. A minor arc of a curve of constant curvature $1/R$ may be assumed to be an arc of a circle of radius R . Assume the entire curve as consisting of a very great number of such minor arcs with two neighbouring arcs partially overlying each other. Two minor arcs of a circle of the same radius partially overlying each other make up a new arc of the same radius. Thus each pair of adjoining minor arcs into which the curve is divided forms an arc of a circle of radius R . Reasoning in this way we shall make certain that each three, four, five, etc. successive minor arcs form an arc of the circle of radius R . Hence, the entire curve forms an arc of the circle of radius R . If we have a closed curve with constant curvature R , it is simply a circle of radius R .

Assume a closed curve q to be the shortest of all the curves bounding a given area F . Suppose it is not a circle, i.e. it is not characterized by an equal curvature at all of its points.

Suppose that, for example, at points A and B of this curve (Fig. V.5), the curvature is different and is equal, respectively, to

$$\frac{1}{R_1} \text{ and } \frac{1}{R_2}$$

where

$$R_1 \neq R_2$$

To make it more definite, assume that

$$\frac{1}{R_1} < \frac{1}{R_2}$$

Consider two small arcs \widehat{CD} and $\widehat{C_1D_1}$ of curve q , the arcs containing points A and B . Replace arc \widehat{CD} with close arc $\widehat{CA'D}$, and arc $\widehat{C_1D_1}$ with close arc $\widehat{C_1B'D_1}$. Denote the area bounded by \widehat{CD} and $\widehat{CA'D}$ by ΔF_1 , and the area bounded by $\widehat{C_1D_1}$ and $\widehat{C_1B'D_1}$ by ΔF_2 . By virtue of formula (V.7) and owing to the replacement of arc \widehat{CD} with arc $\widehat{CA'D}$, the length of curve q will gain an increment equal to (or equivalent to) $(1/R_1) \Delta F_1$, and, owing to the replacement of arc $\widehat{C_1D_1}$ with arc $\widehat{C_1B'D_1}$, the length of curve q will gain an increment equal to $(1/R_2) \Delta F_2$. The total increment of the area bounded by q is equal to $\Delta F_1 + \Delta F_2$, and the increment of the length is equal to (or equivalent to)

$$\frac{1}{R_1} \Delta F_1 + \frac{1}{R_2} \Delta F_2$$

Now select arcs $\widehat{CA'D}$ and $\widehat{C_1B'D_1}$ so that ΔF_1 and ΔF_2 are equal in absolute value and opposite in sign and so that $\Delta F_1 > 0$ and $\Delta F_2 = -\Delta F_1 < 0$. Then the increment of area $\Delta F_1 + \Delta F_2 = 0$, i.e. the area has not changed with the change of curve q . The increment of the length of q is equal to (or equivalent to)

$$\Delta F_1 \left(\frac{1}{R_1} - \frac{1}{R_2} \right)$$

and since

$$\frac{1}{R_1} < \frac{1}{R_2}$$

then

$$\Delta F_1 \left(\frac{1}{R_1} - \frac{1}{R_2} \right) < 0$$

So the increment of the length of q is negative. Curve q has passed into another curve q_1 of a smaller length bounding the same area. Thus q is not a curve of the minimum length among the other curves bounding the given area.

We can draw a conclusion from the above that *the curve of the minimum length among the curves bounding a given area is a circle.**

4. An isoperimetric problem on a surface. A similar problem may be considered on a surface as well in which case the geodesic curvature $\Gamma = \sin \varphi/R$ replaces the space curvature. For example, if minor arc \widehat{CD} of curve q of the geodesic curvature $\Gamma = \sin \varphi/R$ is replaced by close arc $\widehat{CA'D}$, and the area contained between \widehat{CD} and $\widehat{CA'D}$ is equal to ΔF , increment Δl of the length, i.e. the increment obtained by the curve when arc $\widehat{CA'D}$ replaces \widehat{CD} , is expressed by

$$\Delta l = \Delta F \frac{\sin \varphi}{R} = \Gamma \Delta F$$

Repeating the proof of the previous theorem and replacing everywhere the curvature with the geodesic curvature, we shall have the following theorem.

Of all closed curves on a surface, i.e. curves bounding a given area, the curve with a constant geodesic curvature has the minimum length (on a spherical surface such lines are the great and small circles).

NOTE. On a spherical surface as well as on a plane a curve of constant geodesic curvature is the geodesic circle.

On other surfaces curves of constant geodesic curvature generally are not geodesic circles.

* A number of other proofs of the isoperimetric properties of a circle are given in D. A. Kryzhanovsky's book *Izoperimetry, maksimal'nye i minimal'nye svoystva geometricheskikh figur v obshchedostupnom izlozhenii* (Isoperimeter and Maximum and Minimum Properties of Geometric Figures in a Popular Interpretation). Moscow-Leningrad, ONTI (1938). See also L. A. Lyusternik, *Vypuklye tela* (Convex Bodies), Second edition. Moscow-Leningrad, Gostekhizdat (1941).

CHAPTER VI

Fermat's Principle and Its Corollaries

VI.1. Fermat's Principle

1. Fermat's principle. The considered problems are very close to problems of geometrical optics associated with the so-called *Fermat's principle*.

Consider a flat optical medium with the velocity of light $v = v(x, y) = v(A)$ determined at each point $A(x, y)$. The medium is called *homogeneous* if the velocity of light is the same at all of its points.

Time $T(q)$ required to pass along curve q with the velocity of light is called the *optical length of curve q* .

In a homogeneous optical medium, where the velocity of light is v , the optical length $T(q)$ of curve q is proportional to its usual length $l(q)$ and is equal to

$$T(q) = \frac{1}{v} l(q)$$

Fermat's principle states: *in an optical medium the light passes from point A to point B along the line with the minimum optical length of all the lines connecting A and B .*

It follows that *in a homogeneous optical medium light propagates along straight lines.*

2. Reflection law. Given curve s (Fig. VI.1) in a homogeneous optical medium, the curve reflecting light (mirror). Find line q_0 along which light passes from point A to point B reflecting from curve s . Line q_0 is the shortest of lines q connecting A and B and reflecting from s . This line (see Sec. III.2) is broken line ACB with vertex C on line s , and bisector CD of angle ACB is the normal to curve s at point C .

Angles $\alpha = ACD$ and $\beta = DCB$ of rays AC and CB with normal CD are, respectively, called the *angle of incidence* and the *angle of reflection*. Thus we have come to the *law*

of light reflection of Descartes: the angle of incidence is equal to the angle of reflection.

It follows from the properties of normals to an ellipse and parabola deduced in Sec. III.2 that:

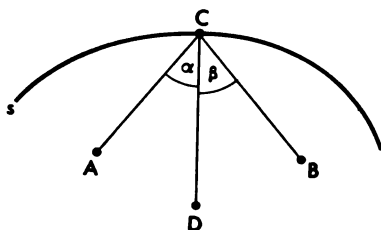


Fig. VI.1

If curve s is an ellipse, the rays passing from focus F of this ellipse, while reflecting, gather in the other focus (Fig. VI.2).

If the curve s is a parabola, the rays passing from the parabola focus, while reflecting, turn into rays parallel to the axis

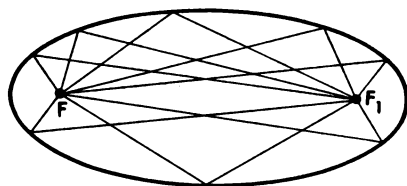


Fig. VI.2

of the parabola and, vice versa, the rays parallel to the axis of the parabola, while reflecting, gather in the parabola focus (Fig. VI.3).

Based on this property of the parabola is the use of mirrors in the form of the paraboloid of revolution (the surface obtained by revolving the parabola about its axis) in search-lights, catoptric telescopes (reflectors), etc.

3. Refraction law. Consider two homogeneous optical media I and II divided by curve s (see Fig. IV.12). The velocity of light in medium I is v_1 and in medium II it is v_2 .

Find the path of light $q_0 = \widehat{AB}$ passing from point A in medium I to point B in medium II .

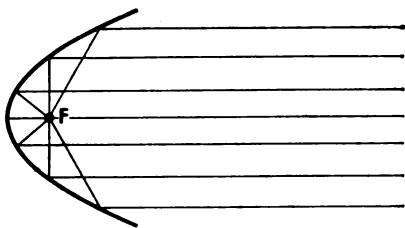


Fig. VI.3

Consider various lines q connecting points A and B and consisting of arcs \widehat{AC} and \widehat{CB} lying, respectively, in media I and II , where C is a point of s . The optical length $T(q)$ of curve q is

$$T(q) = T(\widehat{AC}) + T(\widehat{CB}) = \frac{l(\widehat{AC})}{v_1} + \frac{l(\widehat{CB})}{v_2} \quad (\text{VI.1})$$

Line q_0 is the line of the minimum optical length among all curves q .

Also consider a flexible nonhomogeneous thread q fixed at points A and B , the intermediate point C of the thread sliding along curve s . Tension of part AC of curve q is equal to $T_1 = 1/v_1$ and of part CB , $T_2 = 1/v_2$.

By virtue of formula (IV.7) of Sec. IV.3, the potential energy $V(q)$ is

$$V(q) = \frac{l(\widehat{AC})}{v_1} + \frac{l(\widehat{CB})}{v_2} \quad (\text{VI.2})$$

Comparing formulas (VI.1) and (VI.2), we have

$$T(q) = V(q)$$

The potential energy of thread q coincides with its optical length. So q_0 , i.e. the line of the minimum optical length among curves q , is the line of the minimum potential energy among curves q .

Owing to formula (IV.9) of Sec. IV.3, q_0 is broken line AC_0B . Let α and β be the angles formed by segments AC_0 and C_0B with the tangent LL_1 to curve s at point C_0 . From

formula (IV.9) of Sec. IV.3 it follows that

$$\frac{\cos \alpha}{v_1} = \frac{\cos \beta}{v_2} \quad (\text{VI.3})$$

This gives us the *refraction law of light*. Let α_1 and β_1 be complementary angles to the right angle for α and β , i.e. the angles of segments AC_0 and C_0B with the normal to curve s at point C_0 . Angle α_1 is called the angle of incidence and β_1 the angle of reflection. Formula (VI.3) will be rewritten in the following form:

$$\frac{\sin \alpha_1}{v_1} = \frac{\sin \beta_1}{v_2}$$

VI.2. The Refraction Curve

1. The simplest case. Suppose a plane is divided by straight lines parallel to the x -axis into strips in each of which the velocity of light is constant (Fig. VI.4). Select points A

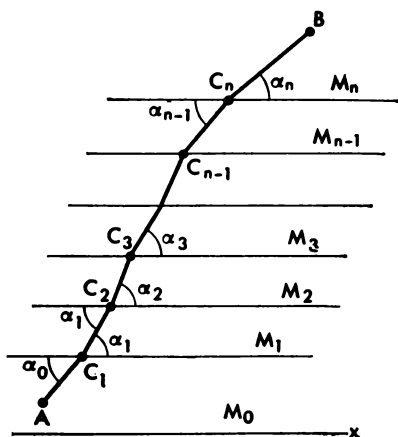


Fig. VI.4

and B lying in different strips. Strip M_0 contains point A , and strip M_n point B . Successively disposed between them are strips M_1, M_2, \dots, M_{n-1} . The velocity of light in strip M_0 is v_0 , in M_1 , v_1 , \dots , in M_n it is equal to v_n . The beam of light passing from point A to point B is in the form

of broken line $AC_1C_2 \dots C_nB$ whose vertices lie on the boundary lines between the strips. Denote the angles between sides $AC_1, C_1C_2, C_2C_3, \dots, C_{n-2}C_{n-1}, C_{n-1}C_n, C_nB$ of this broken line and the straight lines parallel to the x -axis, respectively, by $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_{n-1}, \alpha_n$. By virtue of the refraction law, point C_1 is characterized by the equation

$$\frac{\cos \alpha_0}{v_0} = \frac{\cos \alpha_1}{v_1}$$

point C_2 by

$$\frac{\cos \alpha_1}{v_1} = \frac{\cos \alpha_2}{v_2}$$

etc., and, finally, point C_n by

$$\frac{\cos \alpha_{n-1}}{v_{n-1}} = \frac{\cos \alpha_n}{v_n}$$

Hence, it follows that

$$\frac{\cos \alpha_0}{v_0} = \frac{\cos \alpha_1}{v_1} = \frac{\cos \alpha_2}{v_2} = \dots = \frac{\cos \alpha_{n-1}}{v_{n-1}} = \frac{\cos \alpha_n}{v_n} \quad (\text{VI.4})$$

Denote the common value of these ratios by c , then they can be rewritten in the following form:

$$\frac{\cos \alpha}{v} = c \quad (\text{VI.5})$$

where α = angle of slope of some side of the broken line to the x -axis

v = velocity of light along this side.

The tangent to a broken line at a point of any of its sides is a straight line on which this side lies. Therefore it may be assumed that α in the equation is the angle of slope of the tangent at the point of the broken line to the x -axis, and v the velocity of light at the same point.

2. A refraction curve. Consider an optical medium in which the velocity of light is dependent on the ordinate:

$$v = v(y)$$

where v is a continuous function of y .

The path of light q in such a medium is a line along which

$$\frac{\cos \alpha}{v} = c \quad (\text{VI.6})$$

where v = velocity of light at an arbitrary point C of curve q (Fig. VI.5)
 α = angle between the tangent to q at point C and the x -axis
 c = constant (independent of the selection of point C on the curve).

In order to substantiate equation (VI.6) somewhat, change the distribution of the velocities of light in the medium, viz. divide the latter into narrow strips of width h and consider the velocity of light to be constant in each strip

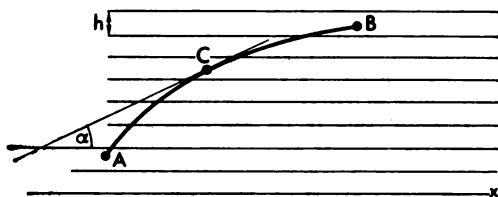


Fig. VI.5

equal, for example, to the velocity of light in the middle of this strip (see Fig. VI.5). Then the path of light from point A to point B will, by virtue of the previous reasoning, be broken line $(AB)_h$ with vertices on the boundary lines between the strips, and equation (VI.6) will hold true for broken line $(AB)_h$ by virtue of the previous considerations. The distribution of velocities has been somewhat changed, and this change is the smaller the narrower are the strips.

As a limiting case, when the width h of strips tends to zero, we shall obtain the initial continuous distribution of the velocities of light. Broken lines $(AB)_h$ will tend to curve q for which conditions (VI.6) will also hold true.

3. Poincaré's model of Lobachevsky geometry. Consider the upper half-plane bounded by the x -axis as an optical medium in which the velocity of light at a point is equal to its ordinate:

$$v = y$$

The beams of light in this medium will be represented by semicircles with the centres at the points of the x -axis (Fig. VI.6).

Let us consider such a semicircle q with the centre at

point O of the x -axis. Let the ordinate at point A of the semicircle be y , and angle ACO of the tangent at this point

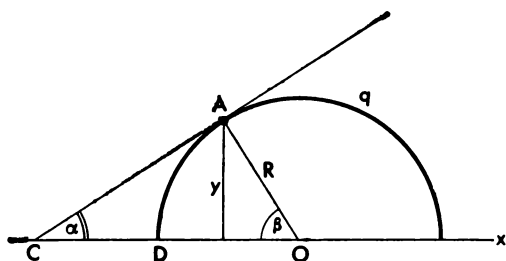


Fig. VI.6

with the x -axis be equal to angle α . If R is the radius of this circle, then

$$y = R \sin \beta$$

where

$$\beta = \angle AOC = \frac{\pi}{2} - \alpha$$

or

$$y = R \cos \alpha$$

i.e.

$$\frac{\cos \alpha}{y} = R$$

So semicircle q satisfies equation (VI.6), i.e. the equation of the beam of light in the medium. The speed tends to zero as we approach the y -axis.

It may be proved that part AD of semicircle q , whose one end lies on the x -axis, has an infinite optical length. For this reason we shall call the points of the x -axis "infinitely remote".

Let us consider semicircles with the centre on the axis as "straight", the optical lengths of the arcs of such semicircles as their "lengths", and the angles at the points of intersection of such straight lines (angles between their tangents) as the "angles" between the straights.

We arrive at a plane geometry that has much in common with many concepts of ordinary plane geometry. Thus only one "straight line" can be drawn through two points (only

a semicircle with the centre on the x -axis can be drawn through two points on a half-plane). The "segment" has the least "length" of all the lines connecting its ends. It is natural to regard as "parallel" two straight lines having a common "infinitely remote point" (i.e. two semicircles touching each other on the x -axis on which their centres lie). Two "straight lines" q_1 and q_2 parallel to q can be drawn through a given point A not lying on "straight line" q

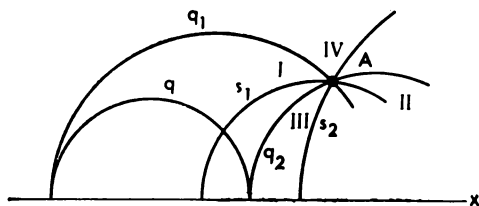


Fig. VI.7

(Fig. VI.7). These straight lines divide the half-plane into four "angles" with vertex A . Straight lines s_1 passing through A and lying in the first pair of vertical angles I and II intersect "straight line" q , whereas all the straight lines s_2 lying in angles III and IV do not intersect q .

We have realized Lobachevsky geometry on a plane, or the so-called Poincaré's model of this geometry.

This model is examined in detail in the book by B. N. Delone*.

VI.3. The Problem of Brachistochrone

1. A cycloid. Let circle K of radius R roll over straight line LL_1 that is assumed to be the x -axis (Fig. VI.8). The motion of the circle is composed of two parts: (1) turning about centre O at an angular speed ω , the linear speed of the points of the circle being $v = R\omega$ and (2) translatory motion parallel to the x -axis at the same speed v . Point A of the circumference circumscribes in the motion a line called the *cycloid*.

* See B. N. Delone *Kratkoe izlozhenie dokazatel'stva neprotivorechivosti planimetrii Lobachevskogo* (Brief Presentation of the Proof of Non-Contradictory Nature of Lobachevsky Planimetry). Moscow, Izd. Akad. Nauk SSSR (1953).

Suppose at moment $t = 0$ point A was on the x -axis (see Fig. VI.8). By the moment t the circle turned through an angle $\beta = t\omega$. The y -coordinate of point A at this moment is equal to

$$y = R(1 - \cos \beta) = 2R \sin^2 \frac{\beta}{2} \quad (\text{VI.7})$$

Find the direction of speed of point A at this moment. This will be the direction of the tangent to the cycloid.

Speed $T_1 = AD_1$ of the translatory motion is equal to v and is parallel to the x -axis. Speed $T_2 = AD_2$ of the motion

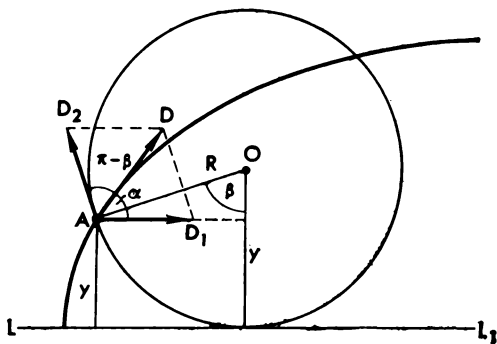


Fig. VI.8

along the circumference is also equal to v and is directed along the tangent to the circumference. Angle D_1AD_2 is equal to $(\pi - \beta)$. Summing up the speeds by the parallelogram law, find the speed of motion of point A along the cycloid. The speed is directed along the bisector of angle D_1AD_2 and forms the angle

$$\frac{1}{2} (\pi - \beta) = \frac{\pi}{2} - \frac{\beta}{2}$$

with the direction of the x -axis (see Fig. VI.5). So the angle α between the tangent to the cycloid at point A and the x -axis is

$$\alpha = \frac{\pi}{2} - \frac{\beta}{2}$$

Therefore

$$\cos \alpha = \sin \frac{\beta}{2} \quad (\text{VI.8})$$

It follows from formulas (VI.7) and (VI.8) that

$$\cos \alpha = \sqrt{\frac{y}{2R}}$$

or

$$\frac{\cos \alpha}{\sqrt{y}} = c \quad (\text{VI.9})$$

Equation (VI.9) links angle α of the slope of the tangent to the cycloid at its point A with the y -coordinate of this point. Inversely, the curve that satisfies this condition is a cycloid.

2. A **brachistochrone**. Suppose A and B are two points, point B assuming to be situated below point A (Fig. VI.9). Connect points A and B by line q . A point moving from A to B along q without the initial speed under gravity will pass curve q in a certain period of time, which is called the *time of fall* along curve q .

Find curve q of the most rapid fall (the *brachistochrone**), a curve connecting points A and B , i.e. a curve for which the time of fall from A to B is the least.

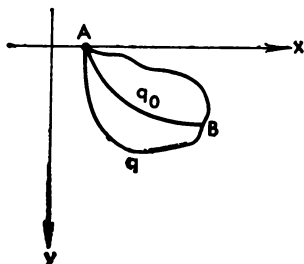


Fig. VI.9

In a vertical plane containing points A and B assume the horizontal straight line on which point A lies as the x -axis and draw the y -axis vertically down. The velocity v of the point moving under gravity without an initial velocity and the y -axis of this point are connected by the relationship

$$v^2 = 2gy$$

or

$$v = \sqrt{2gy} \quad (\text{VI.10})$$

Imagine an optical medium in which the velocity of light v is found from formula (VI.10). The optical length of curve q coincides with the time of fall along the curve. The path of light q_0 from point A to point B is the curve of the least optical length of all the curves connecting points A and B , therefore q_0 coincides with the brachistochrone.

*From the Greek *brachistos*—shortest + *chronos*—time.

For line q_0 equation (VI.6) of Sec. VI.2 is satisfied:

$$\frac{\cos \alpha}{v} = \frac{\cos \alpha}{\sqrt{2g} \sqrt{y}} = c \quad (c = \text{constant})$$

or

$$\frac{\cos \alpha}{\sqrt{y}} = c_1 \quad (c_1 = c\sqrt{2g})$$

Whence, by virtue of the properties of a cycloid outlined above [see formula (VI.9)], the conclusion can be made that the *brachistochrone* is an arc of a cycloid.

VI.4. The Catenary and the Problem of the Minimal Surface of Revolution

1. A **catenary**. A heavy homogeneous chain (or a non-stretchable thread) fixed at two points A and B will, under gravity, assume an equilibrium position along a curve called the *catenary* (Fig. VI.10). (The homogeneity of the chain means that its density ρ is constant; any section of the chain of length h has a mass ρh .)

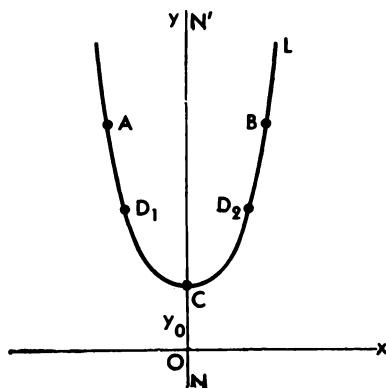


Fig. VI.10

If the chain \widehat{AB} is additionally fixed at points D_1 and D_2 , the position of equilibrium of part $\widehat{D_1 D_2}$ of the chain will not change.

Catenary \widehat{AB} is a continuation of catenary $\widehat{D_1 D_2}$. It may be assumed that the catenary unlimitedly continues on both sides,

or that line \widehat{AB} is part of an *unlimited catenary*.

The lowest point C of the catenary is called its *vertex*. The unlimited catenary is symmetrical with respect to the vertical axis NN' passing through the vertex. Denote this axis by y .

Let us consider the right-hand part \widehat{CL} of the catenary. Denote the ordinate of some point D of the catenary (Fig. VI.11) by y , the angle between the tangent at this point

and the x -axis by α , and the length of arc \widehat{CD} of the catenary by s .

Fix the catenary at points C and D . The force acting on point D is called the *tensile strength* P of the chain at point D and is tangential to the catenary at point D (see Fig. VI.11). Force P_0 acting on point C is tangential to the catenary at this point, i.e. is parallel to the x -axis and is directed to the left.

The resultant T of the gravity forces acting on section \widehat{CD} of the chain is parallel to the y -axis and is directed downwards; the mass of section \widehat{CD} of length s is equal to sp . Whence, T is given by

$$T = gsp \quad (\text{VI.11})$$

where g is the acceleration of gravity.

Force P has a vertical component directed upwards and equal to

$$P \sin \alpha$$

and a horizontal component directed to the right and equal to

$$P \cos \alpha$$

If the catenary solidifies, it will remain in equilibrium. Horizontal forces P_0 and $P \cos \alpha$ and vertical forces T and $P \sin \alpha$ acting on the catenary in opposite directions are balanced. Whence, by virtue of (VI.11),

$$P \sin \alpha = gsp \quad (\text{VI.12})$$

and

$$P \cos \alpha = P_0 \quad (\text{VI.13})$$

Suppose the chain moves along the catenary so that each of its points circumscribes a minor arc of the chain of length h . The chain will occupy the position $\widehat{C'D'}$. Find the work required for such a displacement of the chain.

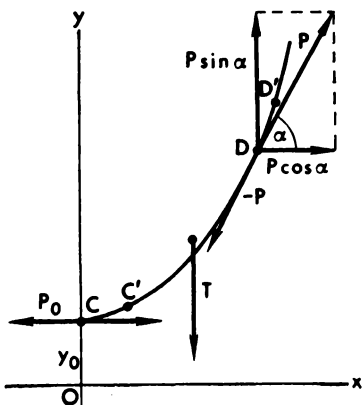


Fig. VI.11

Force P applied to point D will perform the work equal to Ph ; force P_0 applied to point D the work equal to P_0h . Thus the total work spent on the displacement of the chain will be given by

$$W = (P - P_0) h \quad (\text{VI.14})$$

In the former position \widehat{CD} the chain consists of section \widehat{CD} and a small additional section $\widehat{CC'}$. In the new position $\widehat{C'D'}$ the chain consists of the same section $\widehat{C'D}$ and an additional section $\widehat{DD'}$. Both additional sections $\widehat{CC'}$ and $\widehat{DD'}$ are of the same length h and have equal masses ρh , but $\widehat{CC'}$ has a vertical coordinate y_0 , while $\widehat{DD'}$ the y -coordinate. Due to the work spent, instead of the additional section with the y_0 -coordinate another section of the same mass but of the y -coordinate has appeared. Whence we see that the work spent is equal to

$$W = g\rho h (y - y_0) \quad (\text{VI.15})$$

It follows from (VI.14) and (VI.15) that

$$P - P_0 = g\rho (y - y_0)$$

or

$$P - g\rho y = P_0 - g\rho y_0 \quad (\text{VI.16})$$

If the chain is moved parallel to itself in the direction of the y -axis, neither the chain's shape nor its reaction P in different points will change. Move the catenary in the direction of the y -axis so that its initial y_0 -coordinate is equal to

$$y_0 = \frac{1}{g\rho} P_0 \quad (\text{VI.17})$$

Such a position of the catenary is called *standard*. The geometric definition of the standard position of a catenary will be given below.

In this position equation (VI.16) will be simplified and will take the following form:

$$P - \rho g y = 0$$

or

$$y = \frac{1}{\rho g} P \quad (\text{VI.18})$$

The tension at a point of a catenary in standard position is proportional to its ordinate.

It follows from (VI.13) that

$$\frac{1}{\rho g} P \cos \alpha = \frac{1}{\rho g} P_0$$

or, making use of equations (VI.17) and (VI.18),

$$y \cos \alpha = y_0 \quad (\text{VI.19})$$

The latter equation (VI.19) links the ordinate of the catenary point with the angle α between the tangent at this point and the x -axis.

Comparing equation (VI.19) with the equation of the refraction line [see formula (VI.5) in Sec. VI.2] we have

In standard position a catenary is the shape of the light ray in a medium in which the velocity of light v is inversely proportional to the y -coordinate:

$$v = \frac{c}{y}$$

2. The geometric definition of standard position of a catenary. It follows from equations (VI.12) and (VI.18) that

$$\frac{1}{\rho g} P \sin \alpha = s$$

further,

$$s = y \sin \alpha$$

Whence

$$y - s = y (1 - \sin \alpha)$$

Finally, by virtue of (VI.19), we have

$$y - s = y_0 \frac{1 - \sin \alpha}{\cos \alpha}$$

Denote the angle between the tangent to the catenary and the y -axis by $\beta = (\pi/2) - \alpha$. We obtain

$$y - s = y_0 \frac{1 - \cos \beta}{\sin \beta} = y_0 \frac{2 \sin^2 \frac{\beta}{2}}{2 \sin \frac{\beta}{2} \cos \frac{\beta}{2}} = y_0 \tan \frac{\beta}{2} \quad (\text{VI.20})$$

Consider segment DE parallel to the y -axis, directed downwards and equal to length s (to the length of arc \widehat{CD} of the catenary) (Fig. VI.12). If arc \widehat{CD} remains fixed at point D and point C is free, arc \widehat{CD} under gravity will assume a new

equilibrium position, viz. will pass into the vertical segment DE . In short, arc \widehat{CD} of the chain will "fall" into position DE . Segment EE_1 of the vertical line equal to $y - s$ will indicate the distance between end E of the "fallen" part of the chain and the x -axis.

From formula (VI.19) it follows that

$$\sin \beta = \cos \alpha = \frac{y_0}{y} \quad (\text{VI.21})$$

Suppose point D unlimitedly goes upwards along the catenary. Its ordinate will tend to infinity:

$$y \rightarrow \infty$$

Then, by virtue of (VI.21), $\sin \beta$ will tend to zero. But in this

case $\beta \rightarrow 0$ (the angle between the tangent at point D and the y -axis will also tend to zero). But $\tan (\beta/2) \rightarrow 0$ and, by virtue of (VI.20),

$$\lim_{y \rightarrow \infty} (y-s) = 0$$

The distance from end E of the fallen arc \widehat{CD} to the x -axis will tend to zero when end D of this arc tends to infinity.

In standard position of a catenary the x -axis is that horizontal straight line which is unlimitedly approached by the end of the fallen arc \widehat{DE} whose beginning unlimitedly moves away.

These features characterize standard position of a catenary.

3. The minimal surface of revolution. Solve the following problem: of all the plane curves q connecting two given points A and B find the one which in revolving about the x -axis will yield the minimal lateral surface of revolution (Fig. VI.13).

Denote by $V(q)$ the area of the lateral surface of revolution of curve q about the x -axis, and by $T(q)$ the optical length of curve q in a medium where the velocity of light v is taken from the formula

$$v = \frac{1}{2\pi y} \quad (\text{VI.22})$$

Prove the equality of the values

$$V(q) = T(q)$$

Let \widehat{CD} be a minor section of curve q of length h . We must prove that

$$V(\widehat{CD}) = T(\widehat{CD}) \quad (\text{VI.23})$$

Considering \widehat{CD} to be a small straight-line segment and denoting the ordinate of the \widehat{CD} centre of gravity by y , we shall see that the lateral surface of revolution $V(\widehat{CD})$

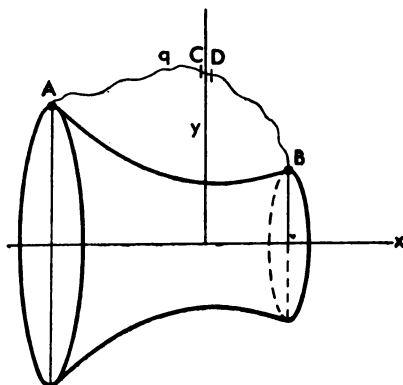


Fig. VI.13

is equal to the lateral surface of the truncated cone with the generatrix equal to h and the radius of the middle section equal to y . Whence

$$V(\widehat{CD}) = 2\pi y h$$

On the other hand, if the velocity of light v in the midpoint of this minor segment (and so approximately for the entire segment) is $1/2\pi y$, its optical length $T(\widehat{CD})$ is given by the relationship

$$T(\widehat{CD}) = \frac{h}{1/2\pi y} = 2\pi y h$$

which is the original equation (VI.23).

It follows from the equality of optical length T and area V of the lateral surface of revolution about the axis (for small sections of curve q) that they are equal for the entire curve q . Therefore, if for q the value of $V(q)$ is the least, the optical length $T(q)$ for the same curve is also the least. By virtue of Fermat's principle, curve q is the light ray in our optical medium connecting points A and B . But in our optical medium the light ray has the shape of the catenary (in standard position).

Thus, of all the curves q connecting points A and B catenary \widehat{AB} (in standard position) has the minimal lateral surface of revolution $V(q)$ about the x -axis.

4. The minimal surfaces. Similarly to the problem of the shortest lines connecting the given points the question may be posed about the minimal surface pulled over a given curve (having the given curve as its boundary), about the so-called *minimal surface*.

If curve r is a plane curve, the part of plane Q bounded by it will be the minimal surface pulled over curve r .

If curve r is not a plane curve, the minimal surface will not be a part of a plane.

Points A and B in revolving about the x -axis form two circles r_1 and r_2 lying in planes perpendicular to this axis and having the centres on the axis. The surface obtained

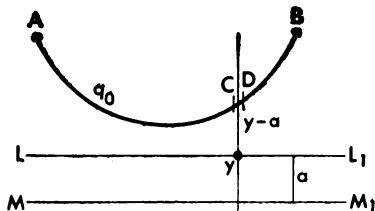


Fig. VI.14

in revolving catenary \widehat{AB} connecting these points is the minimal surface pulled over circles r_1 and r_2 .

5. An isometric problem about the minimal surface of revolution. Here is a more difficult problem: of all the curves of the given length l connecting points

A and B find the one which in revolving about the axis will generate the minimal lateral surface.

Let us consider the axis of revolution LL_1 to be horizontal (Fig. VI.14). Connect points A and B by a chain of the given length l_0 . It will take the form of catenary \widehat{AB} with a given length l_0 . Take for the x -axis a horizontal straight line MM_1 (parallel to the axis of revolution LL_1) such that catenary \widehat{AB} is in standard position with respect to it.

Denote the lateral surface formed by the revolution of curve q about the x -axis (axis MM_1) by $V_1(q)$, the lateral surface formed by the revolution of curve q about axis LL_1 by $V(q)$; $l(q)$ will denote the length of curve q . If a is the distance from axis LL_1 to axis MM_1 , then

$$V(q) = V_1(q) - 2\pi a l(q) \quad (\text{VI.24})$$

Indeed, let \widehat{CD} be a small section of curve q with a length h . If y is the distance from the middle of \widehat{CD} to axis MM_1 , then $y - a$ is its distance to axis LL_1 . Length $l(\widehat{CD}) = h$. Further,

$$V_1(\widehat{CD}) = 2\pi h y, \quad V(\widehat{CD}) = 2\pi h (y - a)$$

Since

$$2\pi h (y - a) = 2\pi h y - 2\pi a h$$

we have

$$V(\widehat{CD}) = V_1(\widehat{CD}) - 2\pi a l(\widehat{CD}) \quad (\text{VI.25})$$

So formula (VI.24) holds true for any small section of curve q . It means that it holds true for the entire curve q as well.

We are interested in curves \bar{q} of length l_0 [$l(q_0) = l_0$] connecting points A and B . For these curves

$$V(\bar{q}) = V_1(\bar{q}) - 2\pi l_0 a$$

i.e. for them the values $V(\bar{q})$ and $V_1(\bar{q})$ differ by a constant value $2\pi l_0 a$. For one and the same curve q_0 these quantities will therefore reach their least values. Catenary q_0 , which is in standard position with respect to the x -axis, will yield the least value of $V_1(q)$ of the lateral surface of revolution about this axis of all the curves connecting A and B , specifically, of curves \bar{q} of length l_0 .

It means that *the same catenary q_0 yields the least values of $V_1(\bar{q})$ of all the curves \bar{q} of length l_0 connecting points A and B .*

This property of the catenary may be proved in a different way.

Consider a set of lines \bar{q} connecting points A and B and having a given length. Each of such lines may be considered as the position of a heavy homogeneous chain of density ρ .

Denote the potential gravitational energy of the chain in position \bar{q} by $U(\bar{q})$. The least value of $U(\bar{q})$ will be for the catenary $\bar{q} = q_0$.

Indeed, by virtue of the Dirichlet principle (see Sec. IV.3), curve q_0 for which $U(\bar{q})$ attains the least value is the position of chain equilibrium.

Let the horizontal straight line MM_1 be the x -axis and assume the density ρ to be equal to 2π . Let this straight line be such for which $U = 0$. If y is the ordinate of the middle of small section \widehat{CD} of the chain of length h (see Fig. VI.14), then

$$U(\widehat{CD}) = \rho h y = 2\pi h y$$

At the same time the lateral surface $V(\widehat{CD})$ of revolution of the same minor arc \widehat{CD} about axis MM_1 (the x -axis) is equal to

$$V(\widehat{CD}) = 2\pi h y$$

Whence

$$U(\widehat{CD}) = V(\widehat{CD})$$

Thus, we have come to the equation

$$U(q) = V(q)$$

Indeed, from the proved equality of U and V for any small part of curve q it follows that they are equal for the entire curve q . Therefore the catenary yielding the least value of $U(\bar{q})$ of all the curves \bar{q} of the given length l connecting points A and B yields the least values for such curves for $V(\bar{q})$ as well.

The value dependent on the curve is called the *functional*. Thus values $l(q)$, $V(q)$, $T(q)$, $U(q)$, etc. are functionals.

Jakob Bernoulli was the first to tackle the following problem.

Of all the curves with a given length find that for which some functional $J(q)$ is maximal or minimal. He named such problems *isoperimetric*. A particular case of such a problem discussed in Sec. V.2 is sometimes called an *isoperimetric problem in the narrow sense*. We have just considered another example of the isoperimetric problem.

VI.5. Interrelation Between Mechanics and Optics

Consider the motion of a point in some flat field (a medium under the action of forces) in which the purely mechanical law of conservation of energy takes place, i.e.

$$U + T = c \quad (\text{VI.26})$$

where $U = U(x, y)$ = potential energy of the moving point
 T = its kinetic energy
 c = total energy (remaining constant during motion).

Assuming the mass of the point equal to unity, we shall have

$$T = \frac{w^2}{2}$$

where w is the speed of the point.

It follows from the latter equation and from (VI.26) that

$$w = \sqrt{2T} = \sqrt{2(c - U)} = \sqrt{2[c - U(x, y)]} \quad (\text{VI.27})$$

Consider various trajectories, i.e. paths circumscribed by the point with the given value of total energy c . It follows from formula (VI.27) that w (the speed of the moving point) is fully characterized by its coordinates x and y , i.e. by its position.

For example, during motion in the gravitational field $U = gy$, where g is the acceleration of gravity and y the upward ordinate, it follows from formula (VI.27) that

$$w = \sqrt{2(c - gy)} = \sqrt{c_1 - c_2 y} \quad (c_1 = 2c, c_2 = 2g) \quad (\text{VI.28})$$

Also consider an optical medium in which the velocity of light v is the inverse of the mechanical speed w :

$$v = v(x, y) = \frac{1}{w(x, y)} \quad (\text{VI.29})$$

The light rays in a medium at a speed of $v = 1/w$ coincide with the trajectories of mechanical motion of a point at a speed of $w = w(x, y)$.

This is the analogy between mechanics and optics established by Hamilton.

It is known, for instance, that in the gravitational field, where the velocity of a point is expressed according to formula (VI.28), the trajectories are in the shape of the parabola; therefore the light rays in an optical medium at the velocity of light $v = 1/\sqrt{c_1 - c_2 y}$ are parabolas.

It is known that the light rays in a medium in which the velocities of light v are proportional to y , $1/y$, and \sqrt{y} are, respectively, semicircles with the centre on the x -axis, catenaries, and cycloids. The same lines are trajectories of a mechanical motion of a point with speeds proportional, respectively, to $1/y$, y , and $1/\sqrt{y}$.

To substantiate this statement, first note that the forces in the field are normal to the equipotential lines, i.e. to the lines of equal potential

$$U(x, y) = \text{const}$$

and are directed towards the smaller magnitudes of potential at each such line [by virtue of formula (VI.27), the

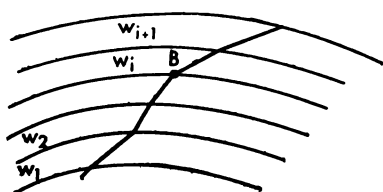


Fig. VI.15

speed $w = w(x, y)$ is also constant]. Draw a set of equipotential lines close to one another. On each such line the speed w is constant and is continuously changing between any two lines. In Fig. VI.15 the lines are denoted by $w_1, w_2, \dots, w_i, w_{i+1}, \dots$, and the

speeds for them are given, respectively, by $w_1, w_2, \dots, w_i, w_{i+1}, \dots$.

Let us replace this motion by another one. Suppose that there is a constant speed w_i between the lines w_i and w_{i+1} , the speed changing abruptly while passing over the line indicated by w_{i+1} . Thus we distort the distribution of speeds, but the nearer the boundary lines (the narrower the space between the lines), the smaller are the changes in speed and the closer is the abrupt distribution of speeds to the original continuous one; the latter may be considered as the limit of abrupt distributions when the width of the strips tends to zero.

In the abrupt distribution of speeds we have force pulses normal to the boundary lines and inducing speed jumps

instead of continuously acting forces (normal to lines $w = \text{const}$).

The forces do not act inside each of the margins and the motion is in a straight line. The trajectories become broken lines with the vertices at the boundary lines. Consider section CBD of such a trajectory (Fig. VI.16). In segment CB the speed is equal to w_{i-1} and is directed along this segment.

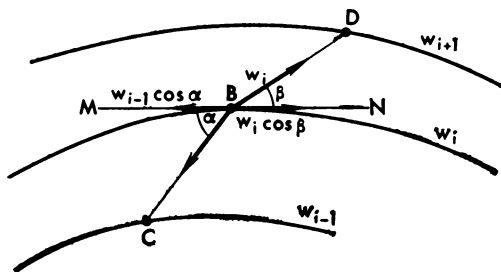


Fig. VI.16

Draw at point B tangent MN to the boundary curve and denote the angles of segments CB and BD with this tangent by α and β . The tangential components of speeds at point B before and after breaking are, respectively, equal to

$$w_{i-1} \cos \alpha \text{ and } w_i \cos \beta$$

Since the force pulse is normal to the boundary curve at point B , it does not change the tangential components, i.e.

$$w_{i-1} \cos \alpha = w_i \cos \beta \quad (\text{VI.30})$$

Formula (VI.30) expresses the refraction law in passing through the boundary line.

Now consider an optical medium for which the velocity of light is the reverse of the speed of mechanical motion $v = 1/w$, i.e. in our adjacent strips I and II the velocity of light is, respectively, $v_{i-1} = 1/w_{i-1}$ and $v_i = 1/w_i$. By virtue of the refraction law at point B we have

$$\frac{\cos \alpha}{v_{i-1}} = \frac{\cos \beta}{v_i}$$

or

$$w_{i-1} \cos \alpha = w_i \cos \beta$$

So the light rays in an optical medium refract similarly to the trajectories in a mechanical medium. Both represent broken lines refracting simultaneously and in the same manner, i.e. the trajectories with speeds w_i in the i th strip coincide with the rays having the velocities of light $v_i = 1/w_i$ in the same strip. Thus the supposition has been proved for abruptly changing media.

In the limiting case when the widths of the strips tend to zero and when a mechanical field is obtained with speeds $w = w(x, y)$ while an optical medium with the velocity of light $v = v(x, y) = 1/w(x, y)$, the coinciding broken trajectories and rays are passed into coinciding curvilinear trajectories and rays.

The interrelation between optics and mechanics indicated by Hamilton is of great importance in contemporary physics.

In conclusion, note that the general methods for solving problems of finding the maximum and the minimum of functionals is the object of investigation of a branch of mathematics called *the calculus of variations*. The fundamentals of this science were laid down by the great mathematicians of the eighteenth century L. Euler and J. Lagrange.

A pamphlet based on lectures read by the author to Moscow University's schools mathematics club.

Deals in an elementary way with a number of variational problems, such as finding the shortest curve uniting two points on a given surface. Suitable for readers with "O" level mathematics.

Contents.

Shortest Lines on Simple Surfaces. Some Properties of Plane and Space Curves and Associated Problems. Geodesic Lines. Problems Associated with the Potential Energy of Stretched Threads. The Isoperimetric Problem. Fermat's Principle and Its Corollaries.

Mir Publishers

Moscow

